Simulation of an Ensemble of Conformational Transitions in a United-Residue Model of Calmodulin

Daniel M. Zuckerman†

Center for Computational Biology & Bioinformatics, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, and Department of Environmental & Occupational Health, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania 15261

Received: October 10, 2003; In Final Form: January 14, 2004

Conformational transitions in proteins provide the mechanism for enacting many critical biological functions at the molecular level. Unfortunately, a full description of the *ensemble* of dynamical pathways in proteins has been out of reach for both experimental and computational techniques. Here, we demonstrate that a simple residue-level computer model of calmodulin's N-terminal lobe yields a statistically converged ensemble of dynamic conformational transitions using ordinary Monte Carlo simulation, without guiding forces, in a few months time on a *single desktop computer*. The ensemble reveals clear pathway heterogeneity, and additionally provides detailed information on intermediate structures. The model uses static structural information from *two* Protein Data Bank files as the sole empirical input, whereas the dynamical results are generated from ordinary simulation.

1. Introduction

The biological functions of many proteins result from their *folded-state* dynamics, and in particular, from conformational changes among metastable states.[1−4] Motor proteins, for instance, perform their vital functions based upon structural transitions, and additionally, such transitions often accompany ligand-binding and catalysis events. Conformational changes occurring in calmodulin, myosin, and adenylate kinase are textbook examples[5] among numerous others. However, little is known about the detailed dynamics of these large-scale transitions. What are the short time scales characterizing the durations of the events? Do metastable intermediate states often occur? How many reaction pathways play an important role?

Calmodulin (CaM), a 148-residue calcium-binding and signaling protein, is an ideal test system because of its modest size and the wealth of experimental data available (e.g., refs 6−25). CaM is a nearly ubiquitous cellular protein, constituting more than 0.1% of all protein in cells.[26] It plays a key role in transducing calcium-mediated signals in processes as diverse as neuronal gene expression, smooth-muscle contraction, and mitosis.[26−28] CaM is toxicologically important as the target of an anthrax exotoxin,[29] and further, engineered forms of CaM are proving useful in monitoring Ca^{2+} localization in living cells (e.g., refs 14, 30, and 31).

Of central interest here, the dumbbell-shaped CaM undergoes large scale conformational changes both within and between its two lobes. Probably the better known change is that occurring between the two lobes in the presence of both bound Ca^{2+} and a target peptide—in this case, the two lobes of CaM close down around the target, leading to a much less extended structure.[5] Yet this docking and compacting process is driven by the intralobe conformational changes (Figure 1) which occur as a result of binding calcium ions. The rearrangement of helices that occurs within each lobe (independently and homologously) is truly dramatic and is the focus of the present study.

CaM has been studied computationally by a number of groups. These prior investigations have employed all-atom force fields, and hence have studied nanosecond-scale dynamic phenonmena. The earliest CaM simulations were performed more than a decade ago and already provided evidence for large-scale interdomain motions mediated by the unwinding of the central helix of the dumbbell.[32,33] Further studies yielded additional data supporting this picture.[34−36] Garcia and co-workers reported intradomain fluctuations in a Ca^{2+}-bound simulation, which were suggestive of the Ca^{2+}-free state,[37] as did Barton et al. in later work.[38] A set of large-scale simulations were performed by Kuczera and co-workers, which provided further insight into the detailed nanosecond dynamics and interactions of calmodulin, with and without bound calcium, as well as in the presence of a target peptide.[36,39−41] Very recent CaM simulations observed significant structural fluctuations in one Ca^{2+} binding loop[42] as well as substantial differences in the stability of the N- and C-terminal lobes.[38,43] Sharp and co-workers performed a careful comparison of simulation and NMR data for a CaM−peptide complex, estimating entropic effects.[44]

The present work differs substantially from previous studies. Our goal is to perform a *complete dynamical simulation* of calmoulin's structural transition—i.e., to capture representative examples of every type of fluctuation and transition pathway—using an inexpensive model. Though we are unaware of any previous report of an ensemble description of protein conformational transition pathways (of the activated transition between two stable folded states), the importance of such a statistical description has long been recognized (e.g., refs 1 and 45), especially for the study of the dynamics of protein folding (e.g., refs 46−52). We note, however, that standard protein folding approaches cannot be applied directly because, in general, the folding models will not enforce the stability of two distinct structures of a single molecule.

Many approaches have been developed for the general problem of determining reaction pathways, including "quasi-static", "quasi-dynamic", and "ensemble" approaches. We term "quasi-static" those approaches yielding a single, presumably optimal pathway (e.g., refs 53−61). The Database of Macromolecular Movements (http://molmovdb.mbb.yale.edu/molmovdb/) provides a notable compendium of quasi-static path-

† Phone: 412-648-3335. Fax: 412-648-3163. E-mail: dzuckerman@ceoh.pitt.edu.

10.1021/jp0370730 CCC: $27.50
Published on Web 03/18/2004

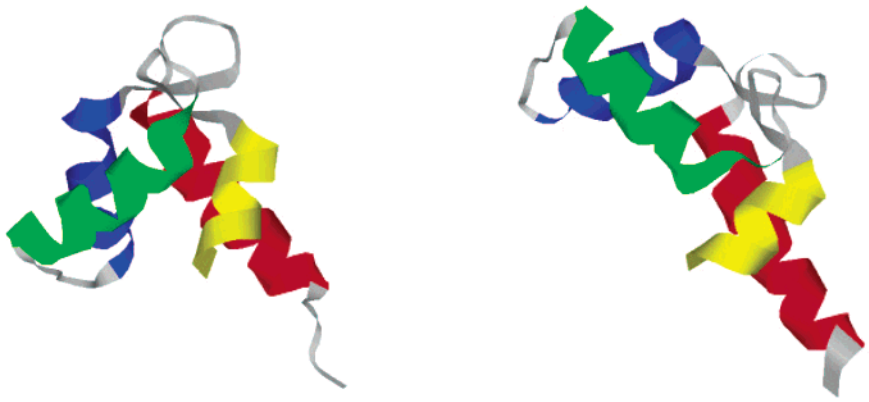

Figure 1. The N-terminal domain of calmodulin (CaM) undergoes a large-scale structural transition upon binding calcium. The left panel is the calcium-free *(*apo*)* structure, PDB code 1cfd, determined by solution NMR, and the right panel depicts the calcium-bound (holo) structure, PDB code 1cll, determined from X-ray diffraction.

ways for protein conformational transitions.[45,62,63] The so-called quasi-static picture, by definition, does not account for inherent thermal fluctuations or the possibility of multiple pathways, though Elber and Shalloway modeled the effect of temperature in a quasi-static picture.[64] "Quasi-dynamic" approaches, like targeted and steered molecular dynamics,[65−68] attempt to include more realistic aspects of the true dynamics; such methods typically generate a small number of biased but potentially dominant dynamical pathways in protein systems.

"Ensemble" approaches attempt to generate suitably distributed sets of reacting trajectories, as does the present study. The pioneering work of Pratt suggesting application of the Metropolis approach to entire path trajectories[69] was applied by Chandler and co-workers to a variety of problems (e.g., refs 70−73). Elber and co-workers developed the stochastic difference equation approach, which generates approximate reactive trajectories without high- frequency motions.[74−76] Woolf and Zuckerman pursued a non-Metropolis ensemble approach,[77−80] as did Mazonka et al.;[81] see also the work of Eastman et al.[82] A notable, independent approach for studying rare events is the "weighted-ensemble Brownian dynamics" of Huber and Kim,[83] which is conceptually related to the earlier work of Harvey and Gabb.[65]

Other previous computational work has pursued goals similar to our own. Grubmiiller and Tavan studied stochastic modeling and memory effects in conformational transitions occurring in a simplified artificial protein.[84] In a recent study, Borovinskiy and Grosberg studied the design of cubic-lattice "protein" models capable of undergoing conformational transitions.[85]

Here we introduce a straightforward, *unbiased* methodology to study large-scale structural transitions in folded proteins, and we use it to perform an initial examination of the structural dynamics in the N-terminal domain of calmodulin. The method builds on existing computational strategies. The first generalizes the work of Gō:[86,87] we use two *static* Protein Data Bank (PDB) structures as input to our model, and study the resulting *dynamics.* Gō models represent proteins in solely a structural way, without reference to the underlying chemistry, aside from size differences of chemical groups. In particular, the Gō model employs a chain of residue "beads", each of which only reacts favorably with other residues that are nearby in the *native* structure. Non-native contacts are penalized. This simple model was developed to study the dynamics of protein folding and it is still widely used for that end.[49,50,88] Note that the Gō model is trivially generalized to other levels of chemical detail, e.g., to all-atom simulation.[50]

Though the present approach is not restricted to simplified models, we use a highly reduced (residue-level) model both because it is easier to implement and because residue-level models have already provided meaningful biomacromolecular results. Reduced models of proteins have often been used in the past, particularly to study protein folding dynamics and thermodynamics (e.g., refs 46, 49, 88−104), "ab initio" protein structure prediction (e.g., refs 105−113), and coarse-grained dynamics (e.g., ref 114); see also a review of structure-based potentials.[115] Moreover, it has been shown conclusively that even models that do not distinguish among atom or amino acid types can capture intermediate scale (i.e., α-carbon) fluctuations and large scale motions (i.e., the slowest modes).[116−119]

Grid-based simulation is also used as a technical device to speed up the simulations, following many precedents,[87,92,93,120−127] particularly the general approach of high coordination lattice studies by Kolinski, Skolnick, and co-workers. Most directly, we employ the simple approach of Panagiotopoulos and Kumar,[124,125] which involves the use of a fine grid (with lattice spacing much less than the particle or atom size) to mimic continuum calculations. Panagiotopoulos and Kumar found that critical and coexistence properties of fluids could be reproduced very precisely with a speed gain of 1−2 orders of magnitude,[124,125] complementing earlier work on proteins.[92,93,122,123] Theoretical studies of relatively fine grid models of fluids, we note, predate protein studies.[128]

Looking ahead, the present, unbiased approach should prove useful in the refinement of the "ensemble" path-sampling methods discussed above. Because the ensemble methods have not been fully vetted in protein systems, the present approach can usefully contribute "gold standard" ensembles of *unbiased* transition trajectories, as a standard for comparison. Further, unbiased ensembles can begin to address a number of methodologically (and biologically) important questions regarding protein transition events: (i) Do metastable intermediates tend to occur? (ii) Is there a very wide variation in the durations of transition events due to the roughness of the energy landscape? (iii) Are multiple pathways typically observed? (iv) How should ensembles of these high-dimensional events be analyzed and categorized?

2. Results

The initial results reported here for fine-grid simulation of a residue-level model of calmodulin are extremely encouraging. The data suggest both a high degree of sampling as well as consistency with experimental results. Only the N-terminal domain was studied—specifically the 72 residues numbered 4−75 in PDB structures 1cfd (apo, unbound) and 1cll (holo,

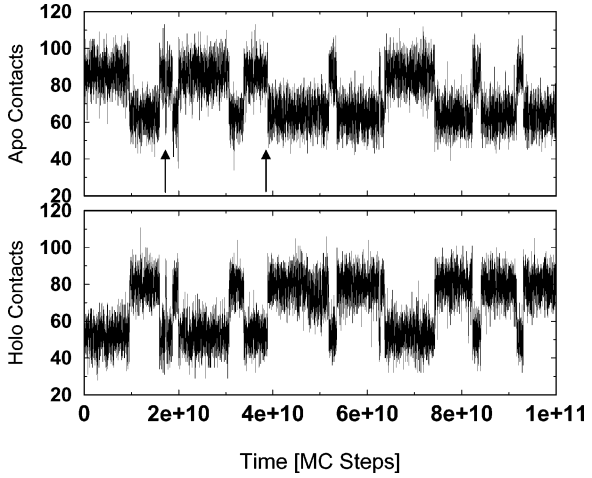

Figure 2. Twelve and one-half days of unbiased single-CPU simulation of the N-terminal domain of calmodulin. More than fifteen transition events are observed. Both apo and holo contacts are depicted to show that transitions occur between the known experimental states. RMS deviations average 2 Å from the experimental structures in each state. The apo contacts are close interresidue contacts (<8 Å) from the simulation which also occur in the 1cfd PDB structure, and holo contacts are those which also occur in the 1cll PDB structure. The arrows indicate transtions depicted in greater detail in subsequent figures.

bound to calcium ions). All simulations were performed at a temperature given by $k_BT = 0.5\epsilon$, which represents the highest temperature at which both apo and holo structures were stable, where ϵ is the square-well depth parameter (see section 3.1).

To summarize the results given below, Figure 3 demonstrates the quite reasonable agreement between the simulation predictions for α-carbon fluctuations and those derived from ensembles of NMR solution structures. This agreement helps to justify the use of a simple model and gives confidence in the data provided by the simulations regarding transitions. Sample transition events and intermediate structures are shown next (Figures 2, 4, 5, and 6). A trace of two helix angles confirms that the transitions are complete and demonstrates the long-time conformational flexibility of the model calmodulin (Figure 5). Finally, the ensemble of trajectories is analyzed for intermediates and for convergence (Figures 6 and 7).

2.1. Equilibrium Backbone Fluctuations. Simulated equilibrium structural fluctuations exhibit good agreement with experimentally determined values, particularly those derived from NMR ensembles, and serve as a reminder of the effectiveness of simple models. Perhaps the most interesting result is that, for the holo case (Figure 3b) where both X-ray and NMR data are available, the simulation exhibits relatively large fluctuations in the loop regions (around residues 23, 42, and 58) in agreement with the NMR results, but somewhat obscured in the X-ray fluctuations. The NMR data also exhibit notable fraying in the N-terminal helix, which is well-captured by the simulation but not by the crystallographic data. These points are particularly notable because the holo-state interactions of the present molecular mechanics potential are constructed from the 1cll X-ray structure (see section 3.1). The phenomenon of solution-state fluctuations deviating from X-ray results is quite consistent with recent NMR studies on both calmodulin[21] and hemoglobin.[129] Further, the average solution state structure of CaM is known to deviate substantially from the crystal structure,[21,22] suggesting that crystallization introduces structural artifacts.[37]

The apo-state results are also encouraging. Figure 3a shows that the simulation results are largely "within the noise" of the different sets of experimental data. These initial simulations do, however, appear to underestimate fluctuations in the region of residues 57−60, the second calcium-binding loop; the helix fraying at the N-term also disagrees somewhat with the NMR data. Note that no apo-state X-ray crystallographic structure is available for calmodulin. We also note that in both the apo and holo states, fluctuations of the ten C-terminal residues, of helix D, were omitted from Figure 3. The simulated helix fraying in this region (similar in magnitude to that of the N-terminal helix) is not appropriately compared to the experimental results because all the NMR studies (except that for the lak8 holo structures) and X-ray studies were performed on full-engthh 148-residue calmodulin. In full-length CaM, of course, helix D is an intermediate helix and constrained from fraying.

In addition to the ensemble-average agreement of simulation and NMR, one can inquire whether individual members of the NMR ensembles occur during the simulation. This question is answered in the affirmative by Figure 5 and the accompanying discussion.

Technical Points Regarding Equilibrium Fluctuations. The NMR fluctuations were calculated on the basis of the variation within each set of structures (ranging from 10 to 30) given in

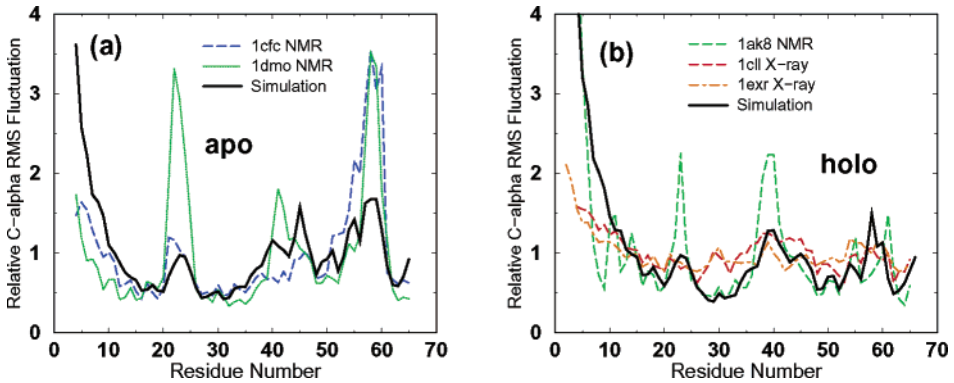

Figure 3. Simulation fluctuations agree qualitatively with NMR results. Fluctuations in α-carbon locations, normalized by the average in every data set, are plotted by residue number for the N-terminal domain. The data are from ensembles of NMR structures, X-ray temperature factors, and simulation results. (a) In the apo state, the simulation data agree reasonably with NMR data, except that the simulation exhibited exaggerated fraying at the N-terminal end and suppressed fluctuations in the calcium-binding loop between helices C and D (residues 57−60). (b) For the holo case, the simulation data again agree qualitatively with principal features of the NMR data, and it is interesting that the increased simulation fluctuations in the N-terminal helix and the three loop regions (around residues 23, 42, and 58) appear to better agree better with the solution-state NMR data than crystallographic X-ray results. Note that holo native-contact interactions built into the simulation are from the X-ray structure 1cll. As discussed in the text, 10 residues from the C-terminal helix D were omitted from this analysis.

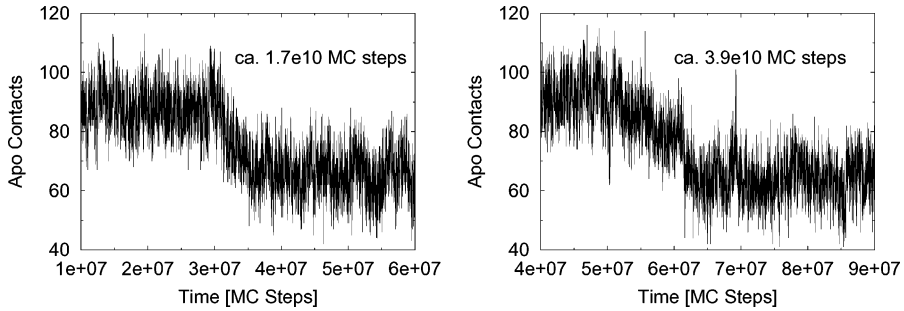

Figure 4. Two sample structural apo-to-holo transitions extracted from the long trajectory of Figure 2. The time points from the longer trajectory are indicated within the plots. The high temporal resolution of the simulation is evident, which complements the long time scales depicted in Figure 2. The transition durations are seen to be of the order of 10^7 MC steps. Note that contact number alone does not dictate which state is occupied: see, e.g., Figure 5 for further structural analysis of the trajectory.

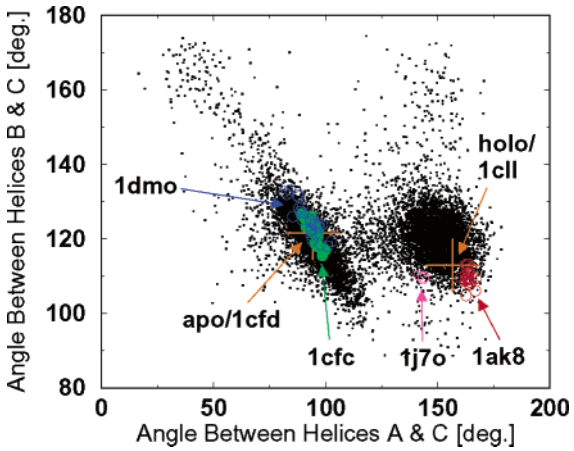

Figure 5. Long trajectory of Figure 2 represented using a density plot in the space of two inter-helix angles. The dots represent configurations, equally spaced in time, generated by the simulation, using Gō interactions to stabilize the PDB structures 1cfd and 1cll. Interhelical angles for experimental structures are also shown, indicated by PDB code. Note that 1cll is the only X-ray structure here, and that multiple data points are present for NMR ensembles. Helix axes were defined to be the vectors associated with the smallest moment of inertia of the residues present in the native helices.

the indicated PDB file, following a number of precedents (e.g., refs 39, 130, and 131). Although such an approach must be used with caution,[132] the NMR ensembles would appear to better represent the long-time scale fluctuations studied here, as contrasted with NMR order parameters that describe picosecond−nanosecond phenomena.[17,133] It should be noted, too, that the process of calculating NMR ensembles is not perfect and is still evolving (e.g., ref 134). The structures were aligned by minimizing the RMS deviation from a reference PDB structure using the TINKER software package[135] and subsequently the RMS fluctuations of the C_α locations were calculated from the sets of aligned structures. The X-ray fluctuations are derived from temperature (B) factors given in the PDB files, via RMS $= \sqrt{3B/8\pi^2}$. Each data set was normalized to give an average of one, in part to obviate the potential artifacts in NMR ensembles.[134]

2.2. Unbiased Structural Transitions in Calmodulin. The main result of the present report is the observation, for the first time in a high-dimensional protein model, of repeated, unbiased, large-scale conformational transitions between distinct stable states; see Figure 2. (By "unbiased," we mean that no guiding forces or time scales were imposed to force the observed

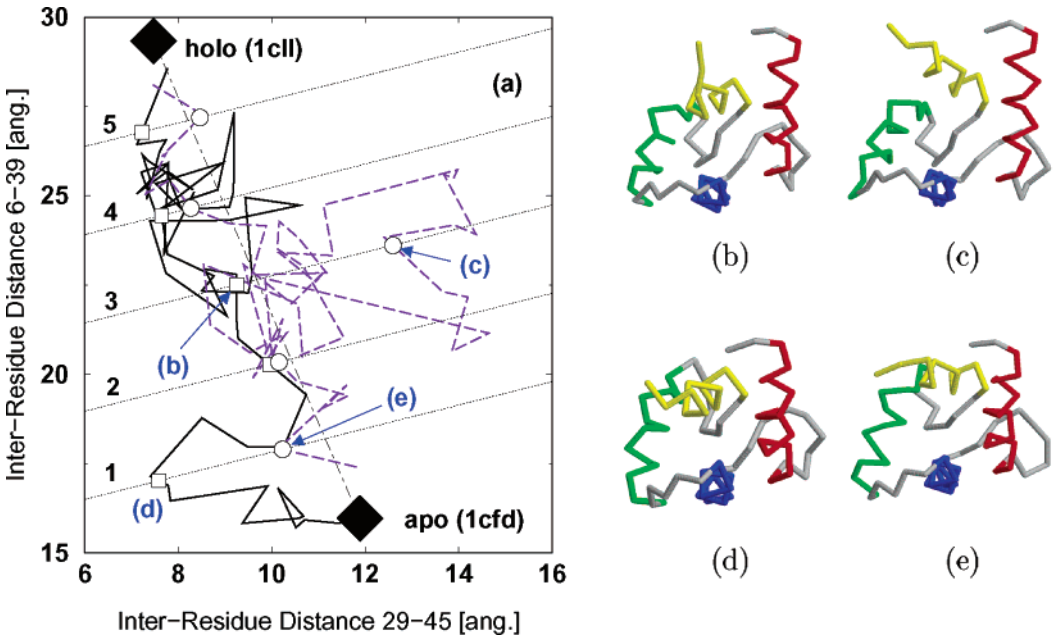

Figure 6. Structural view of the transition events of Figure 4. The main panel (a) shows parametric traces of the two independent transition trajectories with the axes representing distances between the indicated residues. The open circles and squares indicate points where the two trajectories intersect the dotted lines − which are orthogonal to the line between the 1cfd and 1cll structures, although distorted here by scaling. Panels (b)−(e) depict two structures from each path, as labeled in panel (a). Note that residues are numbered from the N-terminal "start" of the peptide, just preceding the red helix: hence the x axis represents the distance between the start of the blue helix and the start of the green; the y axis measures from the start of the red helix to the end of the blue.

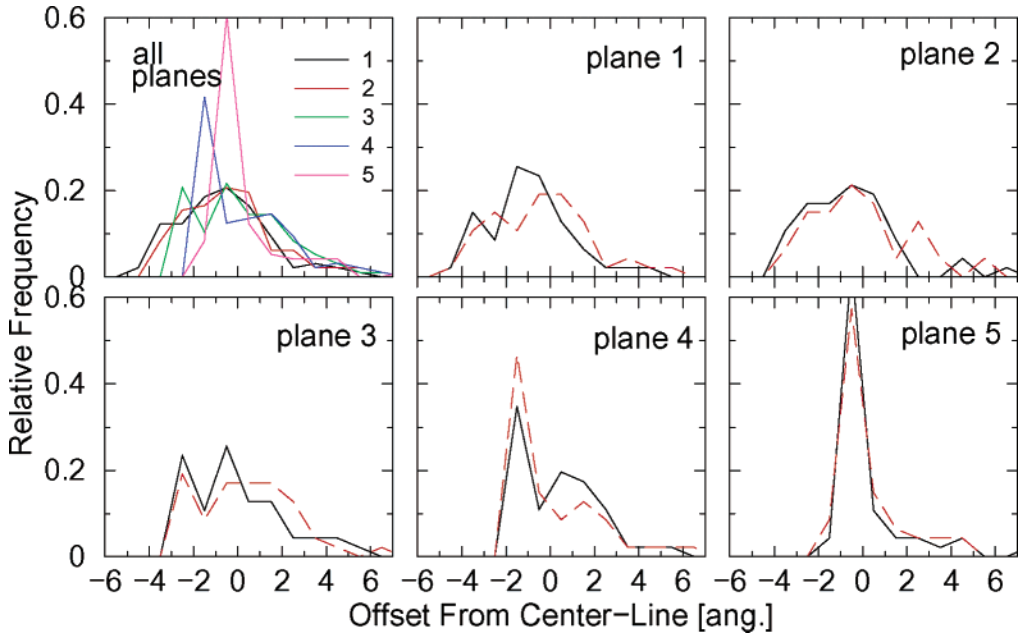

Figure 7. Path distributions show heterogeneity and convergence. Using 94 transition events, distributions are plotted for the crossing points of the planes (lines) labeled 1−5 in Figure 6. The center line defining the zero of the *x* axes is the dot−dashed line connecting the apo and holo structures in Figure 6. The individual plots for the five planes depict the two distributions based on half (47) of the transition events; hence, reasonable convergence may be claimed. For a given trajectory, only the initial crossing points (shown in open symbols for the two paths of Figure 6) were included in the distributions. Note the clear bimodality at planes 3 and 4, and particular sharpness at plane 5.

transitions.) Simulations that can access the details of structural transitions in proteins will be able to augment and guide experimental studies in a significant way.

The two traces of Figure 2 show that not only are transitions made away from the apo state but also those transitions are to the holo state. In each state, the average RMS deviation from the appropriate experimental structure is 2 Å (α-carbons only). The sharpness of the transitions (i.e., the great disparity between the durations of the events themselves and the waiting times between events) is expected for unbiased simulation of an activated two-state process.

Dozens of unbiased transition events were obtained between the apo and holo states of the N-terminal domain of calmodulin (CaM). The simulations possess high-temporal-resolution information on these events, as shown in Figure 4, and can thus reveal structural details of intermediate states and dynamical details regarding lifetimes of intermediate states. We analyze the simulation data at two levels, performing (i) basic checks as to whether the simulations conform to known structural information and (ii) examination of novel simulation results.

Interhelical Angles Confirm Transitions and Demonstrate Access to Large Fluctuations. Figure 5 confirms that the structural transitions observed in the simulations are indeed "real"; that is, the pertinent interhelical angles indeed exhibit transitions *and* the simulations sample the proper apo and holo structures. For comparison, Figure 5 also shows experimental values of the angles from both single-structure PDB files and from multistructure NMR ensemble PDB files.

Figure 5 depicts, furthermore, the extremely broad range of configuration space explored by the simulations, especially by comparison to the NMR ensembles. In this regard, the plot raises an interesting question that will be explored in future studies: Is the range of simulation fluctuations larger than that experienced by CaM in solution or do the present simulations give a reasonable representation of a dynamically fluctuating molecule? Theoretical work by Tang and Dill suggests that, in general, large-scale partial-unfolding fluctuation should be expected in proteins.[136] This issue may be addressed experimentally (e.g., with hydrogen exchange experiments) and by further simulation with more realistic force fields to test the robustness of the present results.

2.3. Pathways and Structural Intermediates. A great strength of the present protocol is its ability to complement long time scale results (i.e., multiple transitions) with fine temporal resolution of individual events. Figure 6 highlights this ability by depicting the trajectories of two example transition events in the space of two revealing coordinates, which represent distances between select helix ends. One in every 10^5 MC steps is shown. Parallel planes (numbered 1−5) provide convenient references for analyzing transition trajectories: the open symbols indicate the points where the trajectories first cross the planes.

The temporal resolution also permits the identification of relatively long-lived intermediates. In Figure 6, one can see that the trajectories seem to spend extra time in the region between planes 4 and 5, as well as near structure b. A more complete analysis of intermediates seems premature with the present simple force field and will be undertaken in future work with more chemically accurate force fields.

2.4. Dispersion and Convergence of the Path Ensemble. We report here, apparently for the first time, on the ensemble properties of a large number of unbiased, full-scale conformational transitions in a nontrivial protein model. Given this wealth of data, many types of analyses are possible, but we focus on one that is visually and physically meaningful.

Transition trajectories are analyzed using the points at which the planes (lines 1−5) of Figure 6 are first crossed by each trajectory. Figure 6 depicts these points for the two example trajectories using open symbols.

The histograms of such crossing points in Figure 7 paint a portrait of diversity, suggesting multiple pathways. The figure highlights both the dispersion in the ensemble, as well as the statistical convergence of the data. All the distributions are skewed and dispersed over eight or more angstroms (where all distances are measured relative to the center line of Figure 6).

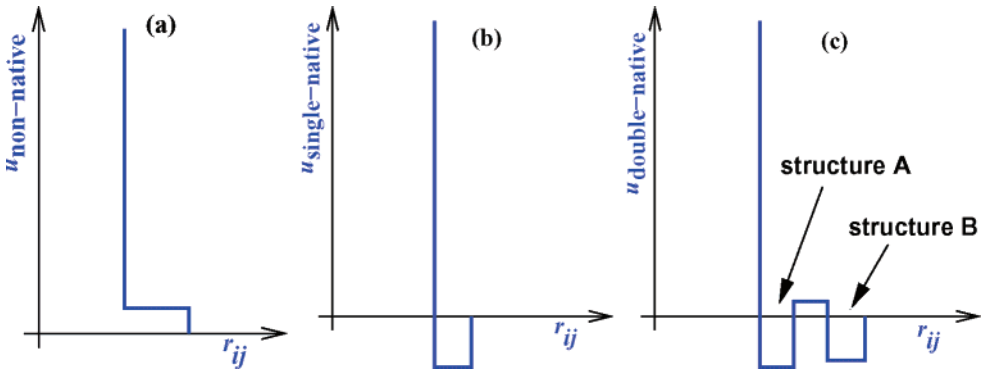

Figure 8. Highly reduced, Gō-like, interresidue potentials used in the simulations. The "non-native" pair potential is used between residues that are not in contact in the holo or the apo structures of calmodulin. A contact is defined by a distance $r_{ij} < 8$ Å between α-carbon atoms of residues *i* and *j* in the Protein Data Bank coordinate files. The "single-native" potential applies for residue pairs in contact in only one of the two structures, whereas "double-native" is for pairs contacting in both.

Two of the distributions (for planes 3 and 4) are clearly bimodal, and the distributions overall show a narrowing trend as the holo state is neared.

Statistical convergence of the crossing-point data is demonstrated in Figure 7 by comparison of the red and black lines in the plots for the individual planes. The two curves in each plot represent distributions resulting from two randomly chosen halves of the data (47 events each). The strong similarity indicates a sufficient ensemble has been obtained to examine structural intermediates using the coordinates of Figure 6.

Event durations also showed a remarkable degree of diversity. Defining a transition trajectory as that continuous part of a trajectory connecting the initial and final states while visiting each only once, the duration may be simply computed as the number of steps in the transition trajectory. The 94 durations studied here varied over nearly 2 orders of magnitude, and the standard deviation (1.2×10^7 MC steps) was substantially larger than the average (7.4×10^6 MC steps). In other words the durations of the transition events vary widely.

The full ensemble of 94 apo → holo transition events requires about two months of single-processor (Intel 2.4 GHz) computer time. Twice as much time was used in fact, because an equal number of "reverse" (holo → apo) events were also generated. However, the "forward" events could readily have been obtained in two months' time using repeated restarts after apo → holo transitions.

Finally, we emphasize that the present results are preliminary in the sense that they reflect a very simple underlying model (section 3.1). Of course, the data should adequately reflect the true behavior of the chosen model; i.e., those phenomena due solely to steric restrictions and native state geometry and attractions. Nevertheless, future work with better force fields will be necessary to assess the ultimate quality of the data.

3. Model and Dynamics

Calmodulin is represented by a simplified, united-residue model, following the spirit of many previous workers.[86,87,116,117,119] The model includes the essential topological features of calmodulin: residue connectivity and size-specific excluded-volume properties, as well as the attractive interactions found in both apo and holo structures. Only the N-terminal domain was included, specifically the 72 residues numbered 4−75 in PDB structures 1cfd (apo, unbound) and Icll (holo, bound to calcium ions). Details are given below.

The simulation dynamics are governed by dynamic Monte Carlo (MC). As described below, this essentially entails ordinary Metropolis MC where only small trial moves are attempted, thus preserving the inherent dynamics as much as possible within the context of MC. Furthermore, residue locations are restricted to the sites of a very fine grid, to speed up computations, with minimal artifacts.

3.1. "Double-Native" Gō Model. The time-independent potential function (Hamiltonian) is constructed using two static, experimentally determined structures: PDB codes 1cfd and 1cll. No other empirical data are used. The model includes covalent, steric, and attractive interactions, with key pair potentials shown in Figure 8. Calcium ions and electrostatic effects are not modeled in this initial study but will be considered in future work.

We introduce a "double-native" generalization of the Gō model;[86] see Figure 8. Traditionally, Gō-like interactions are attractive when two atoms or residues are separated by a small distance near to that found in a "native" structure, but repulsive for non-native contacts. Here, Gō-like attractive square-well interactions occur between residues separated by less than $R_{cut} \equiv 8$ Å in *either* of the native structures, i.e., PDB code 1cfd (structure "A") or 1cll (structure "B"). Following convention, the strengths of attractions and repulsions do not distinguish among residue or atom types, although interaction distances do vary on the basis of native geometries. Thus "chemistry" is included in this pair-specific geometric sense; steric and covalent interactions also model chemistry in a physically reasonable way.

The two-body potential is described in terms of interresidue distances r_{ij} and is of the form

$$U^{tot} = U^{bond} + U^{non\text{-}native} + U^{single\text{-}native} + U^{double\text{-}native} \quad (1)$$

Each pair of residues is classified as non-native, single-native, or double-native, depending on the two distances r_{ij}^A and r_{ij}^B in the native conformations A and B. Specifically, a pair is considered *non-native* if both r_{ij}^A and r_{ij}^B exceed the cutoff distance R_{cut}. A pair is *single-native* if only one of r_{ij}^A and r_{ij}^B exceeds R_{cut}, and *double-native* if both are less than R_{cut}. *This classification is made once at the beginning of the simulation—on the basis of the two PDB structures—and remains in effect throughout.* That is, the Hamiltonian is time-independent.

Bonded interactions, and hence backbone connectivity, were implemented using an infinite square well. Specifically, for a linear chain of *N* beads, the potential is

$$U^{bond}(\{r_{ij}\}) = \sum_{i=1}^{N-1} u^{bond}(r_{i-1,i}) \quad (2)$$

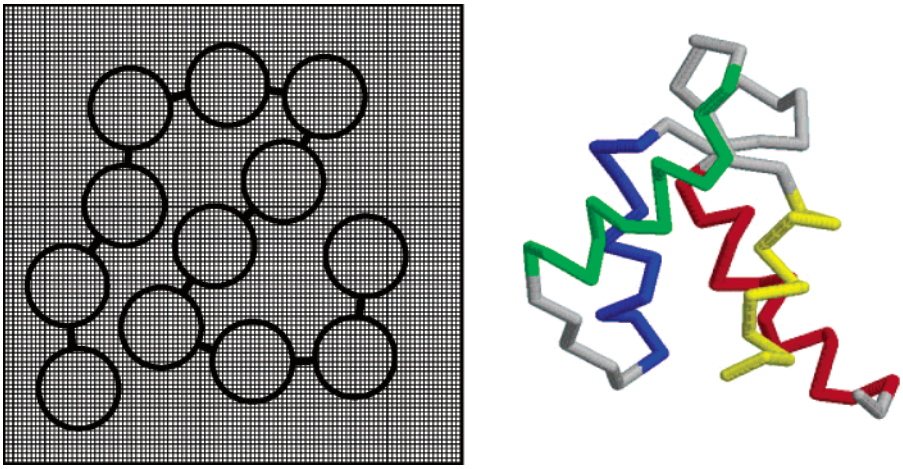

Figure 9. Schematic bead model of a protein on a fine grid, and the residue-level depiction of calmodulin's N-terminal domain. The centers of the residue beads (α-carbon locations) are restricted to grid sites. Because the grid spacing is much less than the size of the beads, the discretized simulation mimics a continuum calculation. Interaction potentials for all possible bead orientations are stored in memory during the simulation, saving potentially costly calculation. The right panel depicts an averaged simulation structure, starting from the configuration of PDB code 1cfd (see Figure 1).

where the pairwise interactions are constructed with reference to interresidue distances in state A (i.e., the apo calmodulin structure 1cfd), denoted r_{ij}^{A}:

$$u^{\mathrm{bond}}(r_{ij}) = \begin{cases} \infty & \text{for } r_{ij} < r_{ij}^{\mathrm{A}}(1-\delta) \\ 0 & r_{ij}^{\mathrm{A}}(1-\delta) \le r_{ij} < r_{ij}^{\mathrm{A}}(1+\delta) \\ \infty & r_{ij} \ge r_{ij}^{\mathrm{A}}(1+\delta \end{cases} \quad (3)$$

All results were generated using the value $\delta = 0.05$.

Non-native interactions are depicted in Figure 8a. They are defined by

$$U^{\mathrm{non-native}}(\{r_{ij}\}) = \sum_{\{i,j\}}^{\text{non-native}} u_{\mathrm{non}}(r_{ij}) \quad (4)$$

where the sum is over pairs classified as non-native according to the time-independent definition given above. The pair interaction potential is given by

$$u_{\mathrm{non}}(r_{ij}) = \begin{cases} \infty & \text{for } r_{ij} < [(\sigma_i + \sigma_j)2/](1-\delta) \\ g_1\epsilon & [(\sigma_i + \sigma_j)/2](1-\delta) \le r_{ij} < R_{\mathrm{cut}} \\ 0 & r_{ij} \ge R_{\mathrm{cut}} \end{cases} \quad (5)$$

The residue hard-core radii σ_i were set to half the distance to the nearest nonbonded residue, with all distances taken between C_α, atoms in the apo calmodulin structure, 1cfd. All the results given above used a shoulder height parameter $g_1 = 0.3$.

Single-native interactions, depicted in Figure 8b, are described by an ordinary square well pair potential. It is defined by

$$U^{\mathrm{single-native}}(\{r_{ij}\}) = \sum_{\{i,j\}}^{\text{single-native}} u_{\mathrm{single}}(r_{ij}) \quad (6)$$

where the sum is over single-native pairs. The pair interaction potential in terms of the native interaction distance $r_{ij}^{\mathrm{X}} < R_{\mathrm{cut}}$ where X is A or B:

$$u_{\mathrm{single}}(r_{ij}) = \begin{cases} \infty & \text{for } r_{ij} < r_{ij}^{\mathrm{X}}(1-\delta) \\ -\epsilon_{\mathrm{X}} & r_{ij}^{\mathrm{X}}(1-\delta) \le r_{ij} < r_{ij}^{\mathrm{X}}(1+\delta) \\ 0 & r_{ij} \ge r_{ij}^{\mathrm{X}}(1+\delta) \end{cases} \quad (7)$$

and $\epsilon_{\mathrm{X}} = \epsilon$ for X = A and $\epsilon_{\mathrm{X}} = g_2\epsilon$ for X = B. The energy value ϵ is simply the unit of energy, by which the temperature scale is defined. On the other hand, g_2, given by 0.95 in the present work, is a parameter that controls the relative lifetimes of states A and B.

Double-native interactions, depicted in Figure 8c, are described by a novel double-square-well pair potential. It is defined by

$$U^{\mathrm{double-native}}(\{r_{ij}\}) = \sum_{\{i,j\}}^{\text{double-native}} u_{\mathrm{double}}(r_{ij}) \quad (8)$$

where the sum is over double-native pairs. Letting the two native distances be ordered such that $r_{ij}^{\mathrm{X}} < r_{ij}^{\mathrm{Y}}$ where X and Y are A or B, the pair potential is

$$u_{\mathrm{double}}(r_{ij}) = \begin{cases} \infty & \text{for } r_{ij} < r_{ij}^{\mathrm{X}}(1-\delta) \\ -\epsilon_X & r_{ij}^{\mathrm{X}}(1-\delta) \le r_{ij} < r_{ij}^{\mathrm{X}}(1-\delta) \\ g_1\epsilon & r_{ij}^{\mathrm{X}}(1+\delta) \le r_{ij} < r_{ij}^{\mathrm{Y}}(1-\delta) \\ -\epsilon_Y & (*)\ r_{ij}^{\mathrm{Y}}(1-\delta) \le r_{ij} < r_{ij}^{\mathrm{Y}}(1+\delta) \\ 0 & r_{ij} \ge r_{ij}^{\mathrm{Y}}(1+\delta) \end{cases} \quad (9)$$

where ϵ_{Y} is defined identically to ϵ_{X}, following eq 7. In the case of overlapping square wells, i.e., $r_{ij}^{\mathrm{X}}(1+\delta) > r_{ij}^{\mathrm{Y}}(1-\delta)$, the middle barrier of height $g_1\epsilon$ simply does not exist and the lower limit of the inequality marked (*) is replaced by $r_{ij}^{\mathrm{X}}(1+\delta)$, by convention. This last choice has little practical impact because the well depths are quite similar ($g_2 = 0.95$, here).

The double-well, attractive interactions depicted in Figure 8c merit further discussion. At first glance, such a potential may seem unnatural. However, amino acids are not spherical, and if one insists on a point-residue model (as here), the only way to represent direct interactions of differing orientations is for the potential to possess more than one well. Indeed, double wells result from statistical analysis of PDB structures.[106,137] A second, perhaps facile, justification comes from the results: as shown above, the double-well potential accomplishes the principal aim of stabilizing two distinct conformational states. A third justification is more pragmatic: when more realistic potentials (e.g., refs 49, 88, 90, 91, 100, 101, 106, and 107) are discretized with the fine-grid approach, presumably the "unnatural" double-

Gō interactions can be reduced to minimal levels required to stabilize the distinct states.

To summarize the potential function, it is a residue-level Gō-like model that "rewards" native contacts found in either of two structures, A or B. Thus the model is built using as input the interresidue distances r_{ij}^{A} and r_{ij}^{B}. Several parameters are also necessary to completely specify the interactions. The parameter δ, here set to 0.05, adjusts the widths of the square wells, whereas the non-native hard-core exclusion distances σ_i are calculated directly from the set of r_{ij}^{A} as described above. The parameter ϵ simply sets the energy scale and so is the unit of the thermal energy scale k_BT. The shoulder and barrier height shown in Figure 8a,c, are both set to $g_1\epsilon$, with the value $g_1 = 0.3$ used in this study. The relative lifetime of state B is governed by the ratio of the depth of the structure B square wells compared to that for A, namely g_2, set to 0.95 here (to ensure nearly equal lifetimes for the apo and holo states). The arbitrary choice of equal lifetimes reflects a desire for optimal computational efficiency, as well as recognition that the simple level of modeling is not expected to accurately capture the relative lifetimes of the states. The temperature throughout was set to $k_BT = 0.5\epsilon$.

3.2. Fine-Grid Simulation. Fine-grid simulation restricts particles (here, residues) to discrete sites on a very fine lattice, permitting the storage in memory of all required information. Thus the approach substitutes rapid "look-ups" for typically quite expensive calculations of forces and/or energies. The general grid/lattice strategy has been in use for decades in protein systems.[87,92,93,120−123] Our implementation for proteins is strongly motivated by high-coordination lattice models of proteins, notably those of Kolinski and Skolnick;[92,93] see also the work of Olson and co-workers.[122,123] The approach we follow directly was developed by Panagiotopoulos and Kumar for liquids simulation[124,125] and has also been applied to polymers.[126,127]

We emphasize that, here, the lattice is simply a device for rapidly implementing *arbitrary continuum models.* Thus *the grid spacing is much smaller than the residue size*, in contrast to conventional lattice models. The fine spacing permits calculations that essentially mimic continuum approaches with minimal artifacts. Indeed, recent work suggests that, for polymer systems, a particle-to-grid size ratio of 17 or 18 is sufficient to prevent artifacts with atomically detailed potentials, whereas 12 suffices for united-atom models.[126,127] In the present work, a conservative ratio exceeding 30 is employed, primarily to implement dynamic Monte Carlo dynamics (see below) with a sufficiently large acceptance ratio. The grid spacing is 0.13 Å.

We point out that the fine-grid approach is readily adaptable to models more complex than that employed in this initial implementation (see section 3.1). Note first that a smooth potential (e.g., Lennard-Jones) is trivially discretized,[125] so one is not restricted to square-well potentials. Multibody, angular terms may also be treated. For instance, bond-angle interactions can be stored in memory as arrays indexed by three discrete distances. Furthermore, in a model with reasonably constant bond lengths, an angle is effectively specified by a single distance. Similar considerations apply to torsion angles.

3.3. Dynamic Metropolis Monte Carlo. In this initial implementation, we have chosen to use dynamic Monte Carlo (DMC) dynamics, which is a common choice for simulations of protein folding dynamics (e.g., refs 49, 50, and 88). Though DMC dynamics were chosen for ease of implementation, they may be justified on two grounds. (i) Because only small, local trial moves will be considered (see below) and accepted with a Boltzmann-factor-preserving probability, DMC may be considered a variant of overdamped Brownian dynamics, which has a well-understood physical basis.[138] (ii) Due to the highly simplified nature of the model used here (see section 3.1), one can expect the simulated dynamics, of any kind, only to give a qualitative or semiquantitative picture. This level of description should not be sacrificed by the Brownian-like dynamics embodied in DMC.

The DMC protocol employed here was designed to produce dynamics as local and realistic as possible. No elaborate schemes were used to artificially accelerate the sampling of configuration space. Specifically, at every time step, a randomly chosen residue makes a trial move with uniform probability to one of the 26 lattice sites on the surface of the $3 \times 3 \times 3$ cube centered at its present location. In other words, trial moves are uniform in all directions, so that favorable directions are determined solely by acceptance/rejection. A trial move is accepted or rejected according to the usual Metropolis criterion: accept if and only if the energy decreases or a random number R_u, chosen uniformly from the interval $0 < R_u \leq 1$ is less than $\exp(-\Delta U/k_BT)$, where ΔU is the change in the total energy (Hamiltonian) of the system, k_B is Boltzmann's constant, and T is the temperature. The temperature throughout was set to $k_BT = 0.5\epsilon$.

3.4. Hardware, Software, and Computer Time. The results reported below are based on a computer program written in the C language and compiled with the "gcc" compiler on Linux machines. Simulations were run on 2.4 GHz Intel processors, which permit approximately 3×10^8 Monte Carlo steps per hour.

4. Summary and Conclusions

We have introduced an approach for studying the long-time dynamics of large-scale conformational transitions in proteins. The technique was tested on a highly reduced, residue-level model of the N-terminal domain of calmodulin, which undergoes a dramatic rearrangement of its four helices upon the (un)binding of calcium ions. Such rearrangements were observed approximately once per day on a single processor in *unbiased* dynamic Monte Carlo simulation. (Here, "unbiased" means a time-independent force field, and no guiding forces, were used.) The scheme is quite promising because it spans the range of time scales from that for interresidue fluctuations to the "waiting time" between transition events, i.e., the inverse rate, which is observed to exceed 10 μs in NMR experiments on the very similar C-terminal domain.[20]

A number of extremely encouraging results were presented: (i) for the first time, a simulation has generated a full ensemble of unbiased transition pathways for a large-scale conformational transition in a nontrivial protein model, and using "desktop" computer resources; (ii) a range of structural intermediates was readily and quantitatively characterized; and (iii) a novel generalization of the Gō model can easily stabilize multiple distinct, experimentally-determined conformational states in a single time-independent force field.

Despite the apparent simplicity of the model employed here, the simulated structural transitions in calmodulin exhibited highly complex behavior. A wide range of transition pathways and event durations were observed. The transition details observed in the present study cannot immediately be identified with expectations for experimental outcomes because of the simplicity of the model. However, the good agreement between the simulation equilibrium fluctuations and NMR results suggests that the present model is a quite reasonable starting point. The model should indeed capture those aspects of the dynamics governed by its physical structure, i.e., size-specific sterics,

backbone connectivity, and native interactions, in analogy with the success of network-like models in predicting large-scale fluctuations.[116−119]

Though the simplicity of the present model precludes a finely detailed comparison with biochemical studies, future studies with more chemically realistic models (e.g., refs 49, 88, 90, 91, 100−103, 106, 107, and 113—see below) should provide experimentally testable hypotheses. Those results verified in a hierarchy of increasingly accurate models are likely to be correct. Note that the effects on dynamics of varying force fields have recently been addressed by two groups for peptide systems.[139,140] A related issue, which also can be addressed directly in the present scheme, regards the effects of *dynamics type*, e.g., Metropolis, Langevin, molecular dynamics, on transition events.

The approach is readily extendable in important ways. First, it is embarrassingly parallelizable, so an inexpensive, modest-sized Beowulf cluster can readily improve upon the single-processor output by an order of magnitude or more. Second, other models (reduced or atomistic) can be stabilized in experimentally determined structural states using the same Gō-like interactions employed here. Improvements in the chemical accuracy of the model are already underway using a "hybrid-Gō" approach (cf. refs 100, 101, and 141) to reduce the somewhat artificial Gō interactions, via a potential of the form

$$U^{\mathrm{tot}} = \alpha\, U^{2\mathrm{Go}} + (1 - \alpha)\, U^{\mathrm{chem}} \qquad (10)$$

where $U^{2\mathrm{Go}}$ is the double-native Gō potential of section 3.1, U^{chem} is a potential grounded in specific physical/chemical interactions, and $0 \leq \alpha \leq 1$ is a paremeter used for "turning off" $U^{2\mathrm{Go}}$. A final generalizable feature is the fine-grid discretization approach, which simply mimics continuum calculations and lends itself to more complex potentials. Although the addition of particles (e.g., "beads") to any model entails a cost, the use of a more chemically realistic potential for U^{chem} will require minimal additional overhead.

The future determination of reaction pathways and dynamics in a hierarchy of models will also address a question both scientifically and methodologically critical: *What is the minimal model necessary to capture residue-level features of structural transitions in proteins?* From the "scientific" perspective, a minimal model will suggest which interactions govern the transition events, whereas, practically speaking, optimal simulation methods should use that model which is least costly.

Finally, it is appropriate to ask what the unbiased transitions observed in the present study presage for the use of the "ensemble" importance-sampling methods discussed in the Introduction.[69,70,73−75,77,78,81,82] The observation of multiple pathways and widely disparate transition event durations suggests that great care will be required in any biased effort to generate ensembles of transition trajectories. Nevertheless, the present approach should provide critical input for testing biased methods in protein systems, namely, gold-standard unbiased trajectories to be used for comparison. From a practical standpoint, a set of reduced-model transition events may also provide useful starting points to sidestep the potential for trapping associated with the transition-path-sampling approach.

Acknowledgment. I owe a special debt of thanks to Ivet Bahar for her invaluable advice, support, and encouragement. Jeffery Evanseck, Eva Meirovitch, Arun Setty, and Marty Ytreberg provided important suggestions and critical feedback on this work. It is also a pleasure to thank David Deerfield, Ron Elber, William Furey, Carol Hall, Hagai Meirovitch, Robert Swendsen, and Dror Tobi for helpful discussions.

References and Notes

(1) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **1991**, *254*, 1598−1603.

(2) Gerstein, M.; Lesk, A. M.; Chothia, C. Structural mechanisms for domain movements in proteins. *Biochemistry* **1994**, *33*, 6739−6749.

(3) Kay, L. E. Protein dynamics from NMR. *Nature Struct. Biol.* **1998**, July supplement, 513−517.

(4) Cavanagh, J.; Venters, R. A. Protein dynamic studies move to a new time slot. *Nature Struct. Biol.* **2001**, *8*, 912−914.

(5) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*, 5th ed.; Freeman: New York, 2002.

(6) Babu, Y. S.; Bugg, C. E.; Cook, W. J. Structure of calmodulin refined at 2.2 angstroms. *J. Mol. Biol.* **1988**, *204*, 191.

(7) Ikura, M.; Clore, G. M.; Gronenborn, A. M.; Zhu, G.; Klee, C. B.; Bax, A. Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* **1992**, *256*, 632−638.

(8) Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A. Calmodulin structure refined at 1.7 angstroms. *J. Mol. Biol.* **1992**, *228*, 1177.

(9) Meador, W. E.; Means, A. R.; Quiocho, F. A. Modulation of calmodulin plasticity in molecular recognition on the basis of X-ray structures. *Science* **1993**, *262*, 1718−1721.

(10) Finn, B. E.; J. Evenäs, Drakenberg, T.; Waltho, J. P.; Thulin, E.; Forsén, S. Calcium-induced structural changes and domain autonomy in calmodulin. *Nature Struct. Biol.* **1995**, *2*, 758−767.

(11) Zhang, M.; Tanaka, T.; Ikura, M. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nature Struct. Biol.* **1995**, *2*, 758−767.

(12) Kuboniwa, H.; Tjandra, N.; Grzesiek, S.; Ren, H.; Klee, C. B.; Bax, A. Solution structure of calcium-free calmodulin. *Nature Struct. Biol.* **1995**, *2*, 768−776.

(13) Evenäs, J.; Thulin, E.; Malmendal, A.; Forsén, S.; Carlstrbm, G. NMR studies of the E140Q mutant of the carboxy-terminal domain of calmodulin reveal global conformational exchange in Ca-saturated state. *Biochemistry* **1997**, *36*, 3448−3457.

(14) Miyawaki, A.; Llopis, J.; Heim, R.; McCaffery, J. M.; Adams, J. A.; Ikura, M.; Tsien, R. Y. Fluorescent indicators for Ca-2+ based on green fluorescent proteins and calmodulin. *Nature* **1997**, *388*, 882−887.

(15) Nelson, M. R.; Chazin, W. J. An interaction-based analysis of calcium-induced conformational changes in Ca sensor proteins. *Protein Sci.* **1998**, *7*, 270−282.

(16) Nelson, M. R.; Chazin, W. J. Structures of EF-hand Ca-binding proteins, diversity in the organization, packing and response to Ca binding. *BioMetals* **1998**, *11*, 297−318.

(17) Evenas, J.; Forsen, S.; Malmendal, A.; Akke, M. Backbone dynamics and energetics of a calmodulin domain mutant exchanging between closed and open conformations. *J. Mol. Biol.* **1999**, *289*, 603−617.

(18) Evenis, J.; Malmendal, A.; Thulin, E.; Carlstrdm, G.; Forsen, S. NMR studies of the E140Q mutant of the carboxy-terminal domain of calmodulin reveal global conformational exchange in Ca-saturated state. *Biochemistry* **1999**, *37*, 13744−13754.

(19) Lee, A. L.; Kinnear, S. A.; Wand, A. J. Redistribution of side chain entropy upon formation of a calmodulin-peptide complex. *Nature Struct. Biol.* **2000**, *7*, 72−77.

(20) Evenäs, J.; Malmendal, A.; Akke, M. Dynamics of the transition between open and closed conformations in a calmodin C-term domain mutant. *Structure* **2001**, *9*, 185−195.

(21) Chou, J. J.; Li, S.; Klee, C. B.; Bax, A. Solution structure of Ca-calmodulin reveals flexible hand-like properties of its domains. *Nature Struct. Biol.* **2001**, *8*, 990−997.

(22) Akke, M.; Chazin, W. J. An open and shut case. *Nature Struct. Biol.* **2001**, *8*, 910−912.

(23) Lee, A. L.; Wand, A. J. Microscopic origins of entropy, heat capacity and the glass transition in proteins. *Nature* **2001**, *411*, 501−504.

(24) VanScyoc, W. S.; Sorensen, B. R.; Rusinova, E.; Laws, W. R.; Ross, J. B. A.; Shea, M. A. Calcium binding to calmodulin mutants monitored by domain-specific phenylalanine and tyrosine fluoresence. *Biophys. J.* **2002**, *83*, 2767−2780.

(25) Weljie, A. M.; Yamniuk, A. P.; Yoshino, H.; Izumi, Y.; Vogel, H. J. Protein conformational changes studied by diffusion NMR spectroscopy, Application to helix-loop-helix calcium binding proteins. *Protein Sci.* **2003**, *12*, 228−236.

(26) Chin, D.; Means, A. R. Calmodulin, a prototypical calcium sensor. *Trends Cell Biol.* **2000**, *10*, 322−328.

(27) Peterson, B. Z.; DeMaria, C. D.; Yue, D. T. Calmodulin is the Ca sensor for Ca-dependent inactivation of L-type calcium channels. *Neuron* **1999**, *22*, 549−558.

(28) Van Eldik, L.; Watterson, D. M. *Calmodulin and signal transduction*; Academic: San Diego, 1998.

(29) Drum, C. L.; Yan, S.-L.; Bard, J.; Shen, Y.-Q.; Lu, D.; Soelalman, S.; Grabarek, Z.; Bohm, A.; Tang, W.-J. Structural basis for the activation of anthrax adenylyl cyclase exotoxin by calmodulin. *Nature* **2002**, *415*, 396−402.

(30) Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. Circular permutation and receptor insertion within green flourescent protein. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11241−11246.

(31) Truong, K.; Sawano, A.; Mizuno, H.; Hama, H.; Tong, K. I.; Mal, T. K.; Miyawaki, A.; Ikura, M. FRET-based in vivo Ca imaging by a new calmodulin-GFP fusion molecule. *Nature Struct. Biol.* **2001**, *8*, 1069−1073.

(32) Mehler, E. L.; Pascual-Ahuir, J.-L.; Weinstein, H. Structural dynamics of calmodulin and troponin C. *Protein Eng.* **1991**, *4*, 625−637.

(33) Weinstein, H.; Mehler, E. L. Ca-binding and structural dynamics in the functions of calmodulin. *Annu. Rev. Physiol.* **1994**, *56*, 213−236.

(34) van der Spoel, D.; de Groot, B. L.; Hayward, S.; Berendsen, H. J. C.; Vogel, H. J. Bending of the calmodulin central helix, a theoretical study. *Protein Sci.* **1996**, *5*, 2044−2053.

(35) Wriggers, W.; Mehler, E.; Pitici, F.; Weinstein, H.; Schulten, K. Structure and dynamics of calmodulin in solution. *Biophys. J.* **1998**, *74*, 1622−1639.

(36) Yang, C.; Jas, G. Z.; Kuczera, K. Structure and dynamics of calcium-activated calmodulin in solution. *J. Biomol. Struct. Dynam.* **2001**, *19*, 247−271.

(37) Vigil, D.; Gallagher, S. C.; Trewhella, J.; Garcia, A. E. Functional dynamics of the hydrophbic cleft in the N-domain of calmodulin. *Biophys. J.* **2001**, *80*, 2082−2092.

(38) Barton, N. P.; Verma, C. S.; Caves, L. S. D. Inherent flexibility of calmodulin domains, a normal-mode analysis study. *J. Phys. Chem. B* **2002**, *106*, 11036−11040.

(39) Yang, C.; Kuczera, K. Molecular dynamics simulations of calcium-free calmodulin in solution. *J. Biomol. Struct. Dynam.* **2002**, *19*, 801−819.

(40) Jas, G. S.; Kuczera, K. Free-energy simulations of the oxidation of C-terminal methionines in calmodulin. *Proteins* **2002**, *48*, 257−268.

(41) Yang, C.; Kuczera, K. Molecular dynamics simulations of a calmodulin-peptide complex. *J. Biomol. Struct. Dynam.* **2002**, *20*, 179−197.

(42) Likic, V. A.; Strehler, E. E.; Gooley, P. R. Dyanmics of Ca-saturated calmodulin D129N mutant studied by multiple molecular dynamics simulations. *Protein Sci.* **2003**, *12*, 2215−2229.

(43) Kleinjung, J.; Fraternali, F.; Martin, S. R.; Bayley, P. M. Thermal unfolding simulations of apo-calmodulin using leap-dynamics. *Proteins* **2003**, *50*, 648−656.

(44) Prabhu, N. V.; Lee, A. L.; Wand, A. J.; Sharp, K. A. Dynamics and entropy of a calmodulin-peptide complex studied by NMR and molecular dynamics. *Biochemistry* **2003**, *42*, 562−570.

(45) Krebs, W. G.; Gerstein, M. The morph server, a standardized system for analyzing and visualizing macromolecular motions in a database framework. *Nucl. Acids Res.* **2000**, *28*, 1665−1675.

(46) Levitt, M.; Warshel, A. Computer simulation of protein folding. *Nature* **1975**, *253*, 694−698.

(47) Sali, A.; Shakhnovich, Karplus, M. How does a protein fold? *Nature* **1994**, *369*, 248−251.

(48) Takada, S.; Luthey-Schulten, Z.; Wolynes, P. G. Folding dynamics with nonadditive forces, a simulation study of a designed helical protein and a random heteropolymer. *J. Chem. Phys.* **1999**, *110*, 11616−11629.

(49) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and energetic factors, what determines the structural details of the transition state ensemble and "en route" intermediates for protein folding? an investigation for small globular proteins. *J. Mol. Biol.* **2000**, *298*, 937−953.

(50) Shimada, J.; Shakhnovich, E. I. The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11175−11180.

(51) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **2002**, *420*, 102−106.

(52) Wang, J. Statistics, pathways and dynamics of single molecule protein folding. *J. Chem. Phys.* **2003**, *118*, 952−958.

(53) Muiller, K.; Brown, L. D. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theor. Chim. Acta (Berlin)* **1979**, *53*, 75−93.

(54) Berkowitz, M.; Morgan, J. D.; McCammon, J. A.; Northrup, S. H. Diffusion-controlled reactions, A variational formula for the optimum reaction coordinate. *J. Chem. Phys.* **1983**, *79*, 5563−5565.

(55) Janin, J.; Wodak, S. J. Reaction pathway for the quaternary structure change in hemoglobin. *Biopolymers* **1985**, *24*, 509−526.

(56) Czerminski, R.; Elber, R. Reaction path study of conformational transitions and helix formation in a tetrapeptide. *Proc. Nat. Acad. Sci. U.S.A.* **1989**, *86*, 6963−6967.

(57) Gillilan, R. E.; Wilson, K. R. Shadowing, rare events, and rubber bands. a variational Verlet algorithm for molecular dynamics. *J. Chem. Phys.* **1992**, *97*, 1757−1772.

(58) Fischer, S.; Karplus, M. Conjugate peak refinement, an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chem. Phys. Lett.* **1992**, *194*, 252−261.

(59) Sevick, E. M.; Bell, A. T.; Theodorou, D. N. A chain of states method for investigating infrequent event processes occurring in multistate, multidimensional systems. *J. Chem. Phys.* **1993**, *98*, 3196−3212.

(60) Vonrhein, C.; Schlauderer, G. J.; Schulz, G. E. Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. *Structure* **1995**, *3*, 483−490.

(61) Kim, M. K.; Chirikjian, G. S.; Jernigan, R. L. Elastic models of conformational transitions in macromolecules. *J. Mol. Graph. Model.* **2002**, *21*, 151−160.

(62) Gerstein, M.; Krebs, W. A database of macromolecular motions. *Nucl. Acids Res.* **1998**, *26*, 4280−4290.

(63) Echols, N.; Milburn, D.; Gerstein, M. MolMovDB, analysis and visualization of conformational change and structural flexibility. *Nucl. Acids Res.* **2003**, *31*, 478−482.

(64) Elber, R.; Shalloway, D. Temperature-dependent reaction coordinates. *J. Chem. Phys.* **2000**, *112*, 5539−5545.

(65) Harvey, S. C.; Gabb, H. A. Conformational transitions using molecular dynamics with minimum biasing. *Biopolymers* **1993**, *33*, 1167−1172.

(66) Schlitter, J.; Engels, M.; Kruger, P.; Jacoby, E.; Wollmer, A. Targeted molecular-dynamics simulation of conformational change − Application to the T + R transition in insulin. *Mol. Simul.* **1993**, *10*, 291−309.

(67) Diaz, J. F.; Wroblowski, B.; Schlitter, J.; Engelborghs, Y. Calculation of pathways for the conformational transition between the GTP- and GDP-bound states of the Ha-ras-p21 protein, Calculations with explicit solvent simulations and comparison with calculations in a vacuum. *Proteins* **1997**, *28*, 434−451.

(68) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. Molecular dynamics study of unbinding of the Avidin−Biotin complex. *Biophys. J.* **1997**, *72*, 1568−1581.

(69) Pratt, L. R. A statistical method for identifying transition states in high dimensional problems. *J. Chem. Phys.* **1986**, *85*, 5045−5048.

(70) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.* **1998**, *108*, 1964−1977.

(71) Dellago, C.; Bolhuis, P. G.; Chandler, D. Efficient transition path sampling, Application to Lennard-Jones cluster rearrangements. *J. Chem. Phys.* **1998**, *108*, 9236−9245.

(72) Csajka, F. S.; Chandler, D. Transition pathways in a many-body system, Application to hydrogen-bond breaking in water. *J. Chem. Phys.* **1998**, *109*, 1125−1133.

(73) Bolhius, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition path sampling, throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.

(74) Olender, R.; Elber, R. Calculation of classical trajectories with a very large time step, Formalism and numerical examples. *J. Chem. Phys.* **1996**, *105*, 9299−9315.

(75) Elber, R.; Meller, J.; Olender, R. Stochastic path approach to compute atomically detailed trajectories, application to the folding of C peptide. *J. Phys. Chem. B* **1999**, *103*, 899−911.

(76) Ghosh, A.; Elber, R.; Scheraga, H. A. An atomically detailed study of the folding pathways of protein a with the stochastic difference equation. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10394−10398.

(77) Woolf, T. B. Path corrected functionals of stochastic trajectories, towards relative free energy and reaction coordinate calculations. *Chem. Phys. Lett.* **1998**, *289*, 433−441.

(78) Zuckerman, D. M.; Woolf, T. B. Dynamic reaction paths and rates through importance-sampled stochastic dynamics. *J. Chem. Phys.* **1999**, *111*, 9475−9484.

(79) Zuckerman, D. M.; Woolf, T. B. Efficient dynamic importance sampling of rare events in one dimension. *Phys. Rev. E* **2001,** *63*, 016702.

(80) Zuckerman, D. M.; Woolf, T. B. Rapid determination of multiple reaction pathways in molecular systems, The soft ratcheting algorithm. www.arXiv.org, physics/0209098, 2002.

(81) Mazonka, O.; Jarzyński, C.; Blocki, J. Computing probabilities of very rare events for Langevin processes, a new method based on importance sampling. *Nucl. Phys. A* **1998**, *641*, 335−354. Erratum. *Nucl. Phys. A* **1999**, *650*, 499−500.

(82) Eastman, P.; Grønbech-Jensen, N.; Doniach, S. Simulation of protein folding by reaction path annealing. *J. Chem. Phys.* **2001**, 8700−8711.

(83) Huber, G. A.; Kim, S. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J.* **1996**, *70*, 97−110.

(84) Grubmiiller, H.; Tavan, P. Molecular dynamics of conformational substates for a simplified protein model. *J. Chem. Phys.* **1994**, *101*, 5047−5057.

(85) Borovinskiy, A. L.; Grosberg, A. Y. Design of toy proteins capable to rearrange conformations in a mechanical fashion. www.arXiv.org, condmat/0212124 2002.

(86) Ueda, Y.; Taketomi, H.; Gō, N. Studies on protein folding, unfolding and fluctuations by computer simulation. i. the effects of specic amino acid sequences represented by specic inter-unit interactions. *Int. J. Peptide Protein Res.* **1975**, *7*, 445−459.

(87) Ueda, Y.; Taketomi, H.; Gō, N. Studies on protein folding, unfolding and fluctuations by computer simulation. II. a three-dimensional lattice model of lysozyme. *Biopolymers* **1978**, *17*, 1531−1548.

(88) Clementi, C.; Jennings, P. A.; Onuchic, J. N. How native-state topology affects the folding of dihydrofolate reductase and interleukinlbeta. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5871−5876.

(89) Levitt, M. A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* **1976**, *104*, 59−107.

(90) Miyazawa, S.; Jernigan, R. L. Estimation of effective interresidue contact energies from protein crystal structures, quasi-chemical approximation. *Macromolecules* **1985**, *18*, 534−552.

(91) Miyazawa, S.; Jernigan, R. L. Residue−residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* **1996**, *256*, 623−644.

(92) Skolnick, J.; Kolinski, A.; Yaris, R. Monte Carlo simulations of the folding of beta-barrel globular proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 5057−5061.

(93) Kolinski, A.; Skolnick, J. *Lattice models of protein folding, dynamics, and thermodynamics;* Chapman and Hall: New York, 1996.

(94) Wilson, C.; Doniach, S. A computer model to dynamically simulate protein folding. *Proteins* **1989**, *6*, 193−209.

(95) Friedrichs, M. S.; Wolynes, P. G. Toward protein tertiary structure recognition by means of associative memory Hamiltonians. *Science* **1989**, *246*, 371−373.

(96) Friedrichs, M. S.; Goldstein, R. A.; Wolynes, P. G. Generalized protein tertiary structure recognition using associative memory Hamiltonians. *J. Mol. Biol.* **1991**, *222*, 1013−1034.

(97) Honeycutt, J. D.; Thirumalai, D. Metastability of the folded states of globular proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3526−3529.

(98) Guo, Z.; Thirumalai, D.; Honeycutt, J. D. Folding kinetics of proteins, a model study. *J. Chem. Phys.* **1992**, *97*, 525−535.

(99) Zhou, Y.; Hall, C. K.; Karplus, M. First-order−disorder-to-order transition in an isolated homopolymer model. *Phys. Rev. Lett.* **1996**, *77*, 2822−2825.

(100) Zhou, Y.; Karplus, M. Folding thermodynamics of a model three-helix bundle protein. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 14429−14432.

(101) Zhou, Y.; Karplus, M. Folding of a model three-helix bundle protein, a thermodynamic and kinetic analysis. *J. Mol. Biol.* **1999**, *293*, 917−951.

(102) Smith, A. V.; Hall, C. K. Alpha-helix formation, discontinuous molecular dynamics on an intermediate-resolution protein model. *Proteins* **2001**, *44*, 344−360.

(103) Smith, A. V.; Hall, C. K. Assembly of a tetrameric alpha-helical bundle, computer simulations on an intermediate-resolution protein model. *Proteins* **2001**, *44*, 376−391.

(104) Micheletti, C.; Banavar, J. R.; Maritan, A. Conformations of proteins in equilibrium. *Phys. Rev. Lett.* **2001**, *87*, 088102.

(105) Tanaka, S.; Scheraga, H. A. Medium- and long-ranged interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules* **1976**, *9*, 945−950.

(106) Liwo, A.; Oldziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. A united-residue force field for off-lattice protein-structure simulations. functional forms, I.; parameters of long-range side-chain interaction potentials from protein crystal data. *J. Comput. Chem.* **1997**, *18*, 849−873.

(107) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Oldziej, S.; Scheraga, H. A. A united-residue force field for off-lattice protein-structure simulations. II. parametrization of short-range interactions, and determination of weights of energy terms by z-score optimization. *J. Comput. Chem.* **1997**, *18*, 874−887.

(108) Liwo, A. Czaplewski, C.; Pillardy, J.; Scheraga, H. A. Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. *J. Chem. Phys.* **2001**, *115*, 2323−2347.

(109) Monge, A.; Friesner, R. A.; Honig, B. An algorithm to generate low-resolution protein tertiary structures from knowledge of secondary structure. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 5027−5029.

(110) Gunn, J. R.; Monge, A.; Friesner, R. A.; Marshall, C. H. Hierarchical algorithm for computer modeling of protein tertiary structure, folding of myoglobin to 6.2-angstrom resolution. *J. Phys. Chem.* **1994**, *98*, 702−711.

(111) Monge, A.; Lathrop, E. J. P.; Gunn, J. R.; Shenkin, P. S.; Friesner, R. A. Computer modeling of protein folding, conformational and energetic analysis of reduced and detailed protein models. *J. Mol. Biol.* **1995**, *247*, 995−1012.

(112) Tobi, D.; Shafran, G.; Linial, N.; Elber, R. On the design and analysis of protein folding potentials. *Proteins* **2000**, *40*, 71−85.

(113) Srinivasan, R.; Rose, G. D. LINUS, A hierarchic procedure to predict the fold of a protein. *Proteins* **1995**, *22*, 81−99.

(114) Bahar, I.; Erman, B.; Halilglu, T.; Jernigan, R. L. Efficient characterization of collective motions and interresidue correlations in proteins by low-resolution simulations. *Biochemistry* **1997**, *36*, 13512−13523.

(115) Jernigan, R. L.; Bahar, I. Structure-derived potentials and protein simulations. *Curr. Opin. Struct. Biol.* **1996**, *6*, 195−209.

(116) Tirion, M. M. Large amplitude elastic motions in proteins from a single- parameter, atomic analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905−1908,.

(117) Bahar, I.; Atilgan, A. R.; Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic model. *Fold. Design* **1997**, *2*, 173−181.

(118) Bahar, I. Dynanics of proteins and biomolecular complexes, inferring functional motions from structure. *Rev. Chem. Eng.* **1999**, *15*, 319−347.

(119) Hinsen, K. Analysis of domain motions by approximate normal mode calculations. *Proteins* **1998**, *33*, 417−429.

(120) Lau, K. F.; Dill, K. A. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules* **1989**, *22*, 3986−3997.

(121) Covell, D. G.; Jernigan, R. L. Conformations of folded proteins in restricted spaces. *Biochemistry* **1990**, *29*, 3287−3294.

(122) Reva, B. A.; Sanner, M. F.; Olson, A. J.; Finkelstein, A. V. Lattice modeling, accuracy of energy calculations. *J. Comput. Chem.* **1996**, *17*, 1025−1032.

(123) Reva, B. A.; Finkelstein, A. V.; Sanner, M. F.; Olson, A. J. Adjusting potential energy functions for lattice models of chain molecules. *Proteins* **1996**, *25*, 379−388.

(124) Panagiotopoulos, A. Z.; Kumar, S. K. Large lattice discretization effects on the phase coexistence of ionic fluids. *Phys. Rev. Lett.* **1999**, *83*, 2981−2984.

(125) Panagiotopoulos, A. Z. On the equivalence of continuum and lattice models for fluids. *J. Chem. Phys.* **2000**, *112*, 7132−7137.

(126) Indrakanti, A.; Maranas, J. K.; Panagiotopoulos, A. Z.; Kumar, S. K. Quantitative lattice simulations of structure and thermodynamics of macromolecules. *Macromolecules* **2001**, *34*, 8596−8599.

(127) Natarajan, S.; Maranas, J. K. Discretization parameters in fine-grained lattice simulations of linear and branched polymers. *J. Chem. Phys.* **2003**, *118*, 9053−9057.

(128) Hall, C. K.; Stell, G. Phase transitions in two-dimensional lattice gases of hard-core molecules with long-range attractions. *Phys. Rev. A* **1973**, *7*, 1679−1689 and references therein.

(129) Lukin, J. A.; Kontaxis, G.; Simplaceanu, V.; Yuan, Y.; Bax, A.; Ho, C. Quaternary structure of hemoglobin in solution. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 517−520.

(130) Bonvin, A. M. J. J.; Bruinger, A. T. Conformational variability of solution nuclear magnetic resonance structures. *J. Mol. Biol.* **1995**, *250*, 80−83.

(131) Abseher, R.; Horstink, L.; Hilbers, C. W.; Nilges, M. Essential spaces defined by NMR structure ensembles and molecular dynamics simulation show significant overlap. *Proteins* **1998**, *31*, 370−382.

(132) Powers, R.; Clore, G. M.; Garrett, D. S.; Gronenborn, A. M. Relationships between the precision of high-resolution protein NMR structure, solution-order parameters, and crystallographic B factors. *J. Magn. Reson. B* **1993**, *101*, 325−327.

(133) Evans, J. N. S. *Biomolecular NMR Spectroscopy*; Oxford University Press: Oxford, U.K., 1995.

(134) Spronk, C. A. E. M.; Nabuurs, S. B.; Bonvin, A. M. J. J.; Krieger, E.; Vuister, G. W.; Vriend, G. The precision of NMR structure ensembles revisited. *J. Biomol. NMR* **2003**, *25*, 225−234.

(135) Ponder, J. W.; Richards, F. M. An efficient newton-like method for molecular mechanics energy minimization of large molecules. *J. Comput. Chem.* **1987**, *8*, 1016−1026. http://dasher.wustl.edu/tinker/.

(136) Tang, K. E. S.; Kill, K. A. How experiments see fluctuations in native proteins, perspective from an exact model. *Int. J. Quantum Chem.* **1999**, *75*, 147−164.

(137) Tobi, D.; Elber, R. Distance-dependent, pair potential for protein folding, Results from linear optimization. *Proteins* **2000**, *40*, 40−46.

(138) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, U.K., 1987.

(139) Zaman, M. H.; Shen, M.; Berry, R. S.; Freed, K. F. Computer simulation of metenkephalin using explicit atom and united atom potentials, similarities, differences, and suggestions for improvement. *J. Phys. Chem. B* **2003**, *107*, 1685−1691.

(140) Mu, Y.; Kosov, D. S.; Stock, G. Conformational dynamics of trialanine in water. 2. comparison of AMBER, CHARMM, GROMOS, and OPLS force fields to NMR and infrared experiments. *J. Phys. Chem. B* **2003**, *107*, 5064−5073.

(141) Zhou, Y.; Karplus, M. Interpreting the folding kinetics of helical proteins. *Nature* **1997**, *401*, 400−403.