Inherent flexibility and protein function: The open/closed conformational transition in the N-terminal domain of calmodulin

ARTICLE *in* THE JOURNAL OF CHEMICAL PHYSICS · JUNE 2008

Impact Factor: 2.95 · DOI: 10.1063/1.2928634 · Source: PubMed

2 AUTHORS, INCLUDING:

Swarnendu Tripathi
University of Houston
17 PUBLICATIONS **101** CITATIONS

Inherent Flexibility and Protein Function: The Open/Closed Conformational Transition in the N-Terminal Domain of Calmodulin

Swarnendu Tripathi and John J. Portman

Department of Physics, Kent State University, Kent, OH 44240

(Dated: November 4, 2007)

Abstract

The key to understanding a protein's function often lies in its conformational dynamics. We develop a coarse-grained variational model to investigate the interplay between structural transitions, conformational flexibility and function of N-terminal calmodulin (nCaM) domain. In this model, two energy basins corresponding to the "closed" apo conformation and "open" holo conformation of nCaM domain are connected by a uniform interpolation parameter. The resulting detailed transition route from our model is largely consistent with the recently proposed EFβ-scaffold mechanism in EF-hand family proteins. We find that the N-terminal part in calcium binding loops I and II shows higher flexibility than the C-terminal part which form this EFβ-scaffold structure. The structural transition of binding loops I and II are compared in detail. Our model predicts that binding loop II, with higher flexibility and early structural change than binding loop I, dominates the conformational transition in nCaM domain.

INTRODUCTION

Many protein functions fundamentally depend on structural flexibility. Complex conformational transitions, induced by ligand binding for example, are often essential to proteins participating in regulatory networks or enzyme catalysis. More generally, a protein's ability to sample a variety of conformational sub-states implies that proteins have an intrinsic flexibility and mobility that influences their function.[1,2] While experimental measurement can offer direct dynamical information about specific residues, uncovering the detailed mechanisms controlling conformational transitions between two meta-stable states is often elusive. In this paper we present an analytic model that aims to clarify the relationship between main-chain dynamics and the mechanisms controlling conformational transitions of flexible proteins. In particular, we examine the mechanism for the open/closed transition of the N-terminal domain of Calmodulin (nCaM) to explore how calcium binding and target recognition can be understood by changes in the mobility and the degree of partial order of the protein backbone.

Calmodulin (CaM) may be an ideal model system to illustrate how conformational flexibility is a major determinant of biological function. CaM is found in all eucaryotic cells and functions as a multipurpose intracellular Ca^{2+} receptor, mediating many Ca^{2+}-regulated processes. CaM is a small (148 amino acid) dumbbell shaped protein with two domains connected by a flexible linker. Each domain of CaM contains a pair of helix-loop-helix Ca^{2+} -binding motifs called EF-hands (helices A/B and C/D in the N-terminal domain). These two EF-hands are connected by a flexible B/C helix-linker (see Fig. 1). In each domain the four helices of apo-CaM are directed in a somewhat antiparallel fashion giving the domains a relatively compact structure while leaving the Ca^{2+}-binding loops exposed. The conformational change induced by binding Ca^{2+} can be described as a change in EF-hand interhelical angle (between helices A/B and C/D) from nearly antiparallel (apo, closed conformation) to nearly perpendicular (holo, open conformation) orientation. Further this domain opening mechanism in nCaM indicates that binding of Ca^{2+} occurs almost exclusively within EF-hands, not between them.[3] The structural rearrangement from closed to open exposes a large hydrophobic surface rich in Methionine residues responisble for molecular recognition of various cellular targets such as myosin light chain kinase.

The high flexibility of CaM is essential to its function. The flexibility of the central helix

linking the two domains allows the activated domains to simultaneously interact with target peptides. The conformational flexibility of the domains themselves allow for considerable binding promiscuity of target peptides, a property essential to its function as a primary messenger in Ca^{2+} signal transduction.[4,5] While similar in structure and fold, the two domains of CaM are quite different in terms of their flexibility, melting temperatures, and Ca^{2+}-binding affinities.[6,7]

The conformational dynamics of Ca^{2+}-loaded and Ca^{2+}-free CaM are well characterized by solution NMR.[5,8] Site specific internal dynamics monitored by model free order parameters S^2, indicate that the helices of the apo-CaM domains are well-folded on the picosecond to nanosecond timescale, while the Ca^{2+}-binding loops, helix-linker and termini are more flexible.[9] On the other hand, spin-spin relaxation (or transverse auto-relaxation) rates, R_2, indicate that the free and bound forms of the regulatory protein exchange on the millisecond timescale.[10] Akke and coworkers have investigated the rate of conformational exchange between the open and closed conformational substrates of C-terminal CaM (cCaM) domain by NMR ^{15}N spin relaxation experiments.[11] Comparison of exchange rates as a function of Ca^{2+} concentration have established that the conformational exchange in apo-cCaM involves an equilibrium switching between the closed and open states that is independent of Ca^{2+} concentration.[9]

X-ray crystallography temperature factors give additional insight into the conformational freedom and internal flexibility of CaM in the open and closed state. Recently, Grabarek proposed a detailed mechanism of Ca^{2+} driven conformational change in EF-hand proteins based on the analysis of a trapped intermediate X-ray structure of Ca^{2+}-bound CaM mutant.[12] This two-step Ca^{2+}-binding mechanism is based on the hypothesis that Ca^{2+}-binding and the resultant conformational change in all two EF-hand domains is determined by a segment of the structure that remains fixed as the domain opens. This segment, called the EF-hand-β-scaffold, refers to the bond network that connects the two Ca^{2+} ions. It includes the backbone and the two hydrogen bonds formed by the residues in the 8^{th} position of binding loops (Ile27 and Ile63) and the C=O groups of the residues in the 7^{th} position of the binding loops (Thr26 and Thr62).[13] Indeed, in the absence of Ca^{2+}, the N-terminal end of the binding loop is found to be poorly structured and very dynamic from NMR structures[11,14,15] and X-ray temperature factors.[12] Functional distinction between the two ends of the binding loops in the domain opening mechanism is buttressed by the great variability of the amino

acid sequences of the N-terminal ends of the Ca^{2+}-binding loops compared with the more conserved C-terminal ends across a variety of different EF-hand Ca^{2+}-binding proteins.[13]

In this paper, we study the role of flexibility in the conformational transition of CaM through an extension of a coarse-grained variational model developed to characterize protein folding.[16,17,18] This model accommodates two meta-stable folded conformations as minima of the calculated free energy surface. The natural order parameters of this model, discussed in detail in the methods section, is well suited to describe partially ordered ensembles essential to the conformational dynamics of flexible proteins. Transition routes and conformational changes of the protein are determined by constrained minimization of a variational free energy surface parameterized by the degree of localization of each residue about its mean position. The computational time to calculate the transition route for nCaM is on the order of several minutes on a typical single-processor PC.

In addition to extensive experimental work characterizing the inherent flexibility of CaM, our results also benefit from all atom molecular dynamics simulations[19,20] as well as recent coarse-grained simulations inspired by models developed to characterize protein folding.[21,22] Although subject to systematic errors due to approximations, analytic models have the important advantage that the results are free of statistical noise that can obscure simulation results (particularly troublesome when characterizing low probability states).

MODEL AND METHODS

A configuration of a protein is expressed by the N position vectors of the α-carbons of the polypepetide backbone. We are interested in describing transitions between two known structures denoted by $\{\mathbf{r}_i^{N_1}\}$ and $\{\mathbf{r}_i^{N_2}\}$. Partially ordered ensembles of polymer configurations are described by a reference Hamiltonian

$$\mathcal{H}_0/k_{\rm B}T = \frac{3}{2a^2}\sum_{ij}\mathbf{r}_i\Gamma_{ij}\mathbf{r}_j + \frac{3}{2a^2}\sum_i C_i[\mathbf{r}_i - \mathbf{r}_i^N(\alpha_i)]^2. \qquad (1)$$

where T is the temperature and $k_{\rm B}$ is Boltzmann's constant. Here, the first term enforces chain connectivity, in which the connectivity matrix, Γ_{ij}, corresponds to a freely rotating chain with mean bond length $a = 3.8$Åand valance angle between successive bond vectors set to by $\cos\theta = 0.8$.[23] The N variational parameters, $\{C\}$, control the magnitude of the fluctuations about α-carbon position vectors $\mathbf{r}_i^N(\alpha_i) = \alpha_i\mathbf{r}_i^{N_1} + (1-\alpha_i)\mathbf{r}_i^{N_2}$. The N variational

parameters, $\{\alpha\}$ $(0 \leq \alpha_i \leq 1)$, specify residue positions as an interpolation between $\{\mathbf{r}_i^{N_1}\}$ to $\{\mathbf{r}_i^{N_2}\}$.

The Boltzmann weight for a constrained chain described by $\mathcal{H}_0$ is proportional to

$$\omega(\{C\},\{\alpha\}) \propto \exp\left[-\frac{3}{2a^2}\sum_{ij}(\mathbf{r}_i-\mathbf{s}_i)G_{ij}^{-1}(\mathbf{r}_j-\mathbf{s}_j)\right] \tag{2}$$

where G_{ij} denotes the correlations of monomers i and j relative to the mean locations, $G_{ij} = \langle \delta\mathbf{r}_i \cdot \delta\mathbf{r}_j\rangle_0/a^2$ with $\delta\mathbf{r}_i = \mathbf{r}_i - \mathbf{s}_i$. Here, the correlations G_{ij} are given by the matrix inverse $G_{ij}^{-1} = \Gamma_{ij} + C_i\delta_{ij}$, and the mean positions of each monomer $\mathbf{s}_i = \sum_j G_{ij}C_j\mathbf{r}_j^N(\alpha_j)$ interpolate between the coordinates in each native structure,

$$\mathbf{s}_i = \sum_j G_{ij}C_j[\alpha_j\mathbf{r}_j^{N_1} + (1-\alpha_j)\mathbf{r}_j^{N_2}]. \tag{3}$$

The statistical properties of a structural ensemble can be described in terms of the first two moments $\mathbf{s}_i$ and $G_{i,j}$ since $\mathcal{H}_0$ is harmonic.

In this model, the probability for a particular configurational ensemble at temperature T is given by the variational free energy $F(\{C\},\{\alpha\}) = E(\{C\},\{\alpha\}) - TS(\{C\},\{\alpha\})$. Here, $S(\{C\},\{\alpha\})$ is the entropy loss due to localizing the residues around the mean postions

$$S(\{C\},\{\alpha\})/k_{\mathrm{B}} = \frac{3}{2}\log\det G - \frac{3}{2a^2}\sum \mathbf{s}_i\Gamma_{ij}\mathbf{s}_j + \frac{3}{2}\sum C_iG_{ii}. \tag{4}$$

The energy is derived from two-body interactions between native contacts, $E(\{C\},\{\alpha\}) = \sum_{[i,j]}\epsilon_{ij}u_{ij}$, where u_{ij} is the average of the pair potential $u(r_{ij})$ over $\mathcal{H}_0$, and ϵ_{ij} is the strength of a fully formed contact between residues i and j given by Miyazawa-Jernigan interaction parameters.[27] The sum is restricted to a set of contacts determined by pairs of residues in the proximity in each of the meta-stable conformations. The pair potential between two monomers is developed by a sum of three Gaussians $u(r) = \gamma_{\mathrm{s}}e^{-3\beta_{\mathrm{s}}r^2/2a^2} + \gamma_{\mathrm{i}}e^{-3\beta_{\mathrm{i}}r^2/2a^2} - \gamma_{\mathrm{l}}e^{-3\beta_{\mathrm{l}}r^2/2a^2}$. The parameters are chosen so that $u(r)$ has a minimum at $r^* = 1.6a$ with value $u_{ij}(r^*) = -1$ formed by the long-range attractive interactions $(\gamma_{\mathrm{l}} = 6.0, \beta_{\mathrm{l}} = 0.27)$ and intermediate-range repulsive interaction $(\gamma_{\mathrm{i}} = 9.0, \beta_{\mathrm{l}} = 0.54)$ as in Ref. 17. Excluded volume interactions are represented by a short-range repulsive potential with $\beta_{\mathrm{s}} = 3.0$ and γ_{s} is chosen so that each contact has $u_{ij}(0)/\epsilon_0 = 100$, where ϵ_0 is the basic energy unit of the Miyazawa-Jernigan scaled contacts.[27] The energy of a contact between residues i and j

in a partially ordered chain is given by

$$\begin{aligned} \epsilon_{ij} u_{ij} &= \epsilon_{ij} \langle u(r_{ij}) \rangle_0 \\ &= \epsilon_{ij} \sum_{\mathrm{k=(s,i,l)}} \frac{\gamma_{\mathrm{k}}}{(1+\beta_{\mathrm{k}} \delta G_{ij})^{3/2}} \exp\left[-\frac{3}{2a^2} \frac{(\mathbf{s}_i - \mathbf{s}_j)^2}{1+\beta_k \delta G_{ij}} \right]. \end{aligned} \tag{5}$$

In this work, we consider a two-state model in which the contacts are separated into three sets: (*i*) contacts that occur in reference structure (1) only, (*ii*) contacts that occur in reference structure (2) only, and (*iii*) contacts in common from both reference structures. Then, we consider that each contact involved exclusively with only one structure is in equilibrium with energy from the other state (which is zero). That is, we replace the pair energy for contacts in sets (*i*) and (*ii*) according to

$$\epsilon_{ij} u_{ij} = -k_B T \log\left[1 + \exp(-\epsilon_{ij} \langle u(r_{ij}) \rangle_0 / k_B T)\right]. \tag{6}$$

This form is analogous to coupling between conformational basins in folding-inspired molecular dynamics simulation.[24,25,26] Contacts described by Eq. 6 independently switch on or off depending on the conformational density characterized by a set of constraints $\{C, \alpha\}$.

Analysis of the free energy surface parameterized by $\{C, \alpha\}$ follows the program developed to describe folding:[17] the ensemble of structures controlling the transition is characterized by the monomer density at the saddlepoints of the free energy. At this point, we simplify our model and restrict the interpolation parameter α_i to be the same for all residues, $\alpha_i = \alpha_0$ following Kim et al..[28] Then, the numerical problem simplifies to minimizing the free energy with respect to $\{C\}$ rather than finding saddlepoints in $\{C, \alpha\}$.

To explore the nature of conformational dynamics in detail, we apply this model to the N-terminal domain of CaM (nCaM). In particular, we use residues numbered 4-75 of unbound nCaM (apo, 1cfd) and bound nCaM (holo, 1cll) (see Fig. 1). In our model, we have defined closed nCaM (1cfd) as structure (1) and open nCaM (1cll) as structure (2). Thus, the interpolation parameter $\alpha_0 = 1$ corresponds to the closed state, and $\alpha_0 = 0$ corresponds to the open state. The coordinates of the open/closed structure was rotated to minimize the rmsd of α-carbons between the two structures.[29] We note global alignment has the risk of possibly obscuring or averaging out some local structural differences. The temperature T for the open/closed transition is taken to be the folding temperature (T_{f}) of the open (holo, 1cll) structure with $k_{\mathrm{B}} T_{\mathrm{f}} = 2.0$. For comparison, the folding temperature for closed (apo, 1cfd) structure is $k_{\mathrm{B}} T_{\mathrm{f}} = 1.9$.

For a given set of constraints, $\{C, \alpha\}$, the monomer density of a partially ordered ensemble can be characterized by the Gaussian measure of similarity to conformation described by $\{\mathbf{r}_i^{N_1}\}$

$$\begin{aligned}\rho_i^{(1)}[\{C,\alpha\}] &= \left\langle \exp\left[-\frac{3\alpha^N}{2a^2}(\mathbf{r}_i - \mathbf{r}_i^{N_1})^2\right]\right\rangle_0 \\ &= (1+\alpha^N G_{ii})^{-3/2}\exp\left[-\frac{3\alpha^N}{2a^2}\frac{(\mathbf{s}_i - \mathbf{r}_i^{N_1})^2}{1+\alpha^N G_{ii}}\right]. \end{aligned} \tag{7}$$

Similarly, the structural similarity to the conformation described by $\{\mathbf{r}_i^{N_2}\}$ is defined as

$$\rho_i^{(2)}[\{C,\alpha\}] = (1+\alpha^N G_{ii})^{-3/2}\exp\left[-\frac{3\alpha^N}{2a^2}\frac{(\mathbf{s}_i - \mathbf{r}_i^{N_2})^2}{1+\alpha^N G_{ii}}\right]. \tag{8}$$

The structural similarity relative to the native structures given by $\{\rho^{(1)}\}$ and $\{\rho^{(2)}\}$specify local order parameters suitable to describing conformational transitions between metastable states in proteins.

To investigate the detailed main-chain dynamics controlling the structural change in CaM, we characterize the relative similarity to the closed structure along the transition route through the normalized measure

$$\overline{\rho_i}^{(1)}(\alpha_0) = \frac{\rho_i^{(1)}(\alpha_0) - \rho_i^{(1)}(0)}{\rho_i^{(1)}(1) - \rho_i^{(1)}(0)}, \tag{9}$$

where $\rho_i^{(1)}(\alpha_0)$ is the monomer density of the i^{th} residue with respect to the closed conformation (Eq. 7). Similarly, we represent the relative structural similarity to the open conformation as

$$\overline{\rho_i}^{(2)}(\alpha_0) = \frac{\rho_i^{(2)}(\alpha_0) - \rho_i^{(2)}(1)}{\rho_i^{(2)}(0) - \rho_i^{(2)}(1)}, \tag{10}$$

where $\rho_i^{(1)}(\alpha_0)$ is the monomer density of the i^{th} residue with respect to the open conformation (Eq. 8). In the open state, $\overline{\rho_i}^{(1)}(0) = 0$ and $\overline{\rho_i}^{(2)}(0) = 1$, while in the closed state $\overline{\rho_i}^{(1)}(1) = 1$ and $\overline{\rho_i}^{(2)}(1) = 0$. To represent the structural changes more clearly, it is convenient to consider the difference,

$$\Delta\overline{\rho_i}(\alpha_0) = \overline{\rho_i}^{(1)}(\alpha_0) - \overline{\rho_i}^{(2)}(\alpha_0) \tag{11}$$

for each residue. This difference shifts the relative degree of localization to be between $\Delta\overline{\rho_i}(1) = 1$ and $\Delta\overline{\rho_i}(0) = -1$ corresponding to the open and closed conformations, respectively.

RESULTS

Conformational Flexibility and Calcium Binding

The local mean square fluctuations of α-carbon positions (related to the temperature factors from X-ray crystallography) are a natural set of order parameters for the reference Hamiltonian H_0 in our model. This parameter, $B_i = \langle \delta \mathbf{r}_i^2 \rangle_0$, contains information about the degree of structural order and conformational flexibility of each residue. In Fig. 2 we have plotted B_i versus sequence number at different values of α_0, the parameter that controls the uniform interpolation between the open structure ($\alpha_0 = 0$) and the closed structure ($\alpha_0 = 1$). Fig. 3 shows the corresponding 3D structures of nCaM domain with the residues colored according to B_i. Aside from the very flexible ends of two terminal helices A and D, the Ca^{2+}-binding loops and the helix linker possess the highest flexibility. The calculated fluctuations from our model exhibit very good qualitative agreement with X-ray temperature factors[12] and simulation results[21,30] of CaM.

Binding loops. Each EF-hand in CaM coordinates Ca^{2+} through a 12-residue loop: Asp20-Glu31 in loop I and Asp56-Glu67 in loop II. The C-terminal ends of the loops contain a short β-sheet (residues 26-28 in loop I and residues 62-64 in loop II) adjacent the last three residues that are part of the exiting helices B and D, respectively.

As shown in Fig. 2, the loops remain relatively flexible even in the open conformation. The highest flexibility is near the two Glycines in position 4 of the Ca^{2+}-binding loops I (Gly23) and II (Gly59). This invariable Gly residue provides a sharp turn required for the proper geometry of the Ca^{2+}-binding sites.[30,31] The linker between helices B and C is also very mobile, with the highest flexibility near residue Glu45. Taken together, the mobility of the loops and B/C linker indicates that the domain opening depends entirely on a set of inherent dynamics, or "intrinsic plasticity", of CaM.[5]

A closer look at the fluctuations of the Ca^{2+}-binding loops reveals that the N-terminal part of each loop is more flexible than the C-terminal part. This agrees with NMR data characterizing the flexibility of the N-terminal and C-terminal part of loop III and IV of the C-terminal domain.[9,11] In the transition route (from closed $\rightarrow$ open), the N-terminal ends of the loops stiffen gradually. On the other hand, in the C-terminal part of the loops the short β-sheet structure (residues 26-28 in loop I and 62-64 in loop II) remain rigid (see

Fig. 2 and 3). Also the last three residues of the loops (residues 29-31 in loop I/helix B and residues 65-67 in loop II/helix D) remain relatively rigid, stabilized by the exiting helices B and D respectively.[31]

This immobile region, the EF-hand β-scaffold, is central to a recent proposed mechanism for CaM[12] and other EF-hand domains.[13] Fig. 2 shows that residues Thr26 and Ile27 (in β-sheet of loop I) and Thr62 and Ile63 (in β-sheet of loop II) remain very rigid during the domain opening.

It is also interesting to compare the relative flexibility of binding loop I and II. It is clear that binding loop II is more flexible than loop I in the both conformations (see Fig. 2 and 3(a)). In particular, the connection between helix A and the binding loop I is much more rigid than the connection between helix C and the binding loop II. This large difference in flexibility suggests that binding loop II of nCaM is more dominate in the mechanism for the structural transition. A similar mechanism in C-terminal CaM domain was also observed from NMR studies, where the Ca^{2+}-dependent exchange contribution is dominated by binding loop IV with lower S^2 (higher flexibility) than loop III.[9]

Helices B and C and the B/C linker. Fig. 2 and Fig. 3 also shows that the bottom part of helix C (close to B/C helix linker) is very flexible in apo nCaM. Upon opening, the flexibility of helix C decreases significantly. [See the change in color from blue to white (Fig. 3(a)-(c)) at the bottom part (close to B/C helix-linker) of helix C and from white to red at the middle part of helix C.] In contrast, the top part of helix B (close to binding loop I; residues 29–31) becomes more flexible than the bottom part of helix B (close to B/C helix-linker; residues 32–37) during closed to open transition (see Fig. 2). We also note that residues 37–42 of the B/C helix-linker shows significant increase in flexibility during opening of the domain. This change in flexibility of the B/C helix-linker helps facilitate the concerted reorientation of helices B and C during the closed $\rightarrow$ open transition. Similar behavior was also observed in molecular dynamics simulation of CaM[19] for this six-residue (residues 37–42) segment.

Conformational Change and Transition Mechanism

The results discussed in the previous section gives a picture of the closed to open transition with good overall agreement with experiment and simulation results on an isolated apo-

CaM domain. Nevertheless, the analysis has focused primarily on the difference in the magnitude of fluctuations of the two meta-stable states. We now turn our attention to the predicted transition mechanism and qualitative nature of structural changes along the transition route. Such a description includes: along the transition route from closed to open, what structural changes are predicted to occur early/late, and which are predicted to happen gradually/cooperatively. While such details have yet to be revealed directly through measurement, in principle, site-directed mutagenesis experiments can be used to identify kinetically important structural regions of nCaM.

To clarify the transtition route, we introduce a structural order parameter that measures the similarity to the open or closed state, $\Delta\overline{\rho_i}$ given in Eq. 11. This order parameter is defined so that $\Delta\overline{\rho_i} = 1$ corresponds to the closed conformation and $\Delta\overline{\rho_i} = -1$ corresponds to the open conformation of nCaM domain. Fig. 4 illustrates the conformational transition in nCaM domain in terms of $\Delta\overline{\rho_i}$ for each residue. An alternative representation of the same data is shown in Fig. 5; here, the value of $\Delta\overline{\rho_i}$ is represented as colors ranging from red ($\Delta\overline{\rho_i} = -1$) to white ($\Delta\overline{\rho_i} = 0$) to blue ($\Delta\overline{\rho_i} = -1$) superimposed on the interpolated structure for selected values of α_0.

We first notice that an early transition in the binding loops and in the central region of helix C evident in Fig. 4. [See also the gradual change in color from blue to red in the structures of Fig. 5(a)-(d).] We also note the concerted structural change of parts of helices B and C and flexible B/C helix-linker (residues 31–49). In particular, the flexible B/C helix-linker (residues 38-44) in Fig. 4 exhibits a cooperative transition. Residue Gln41 which located in this linker region is highly mobile according to NMR data.[14,15] The change in color from red to blue in the B/C helix linker in Fig. 5(a) and (b) indicates that the structural transition of the N-terminal part (close to helix B) of this linker occurs earlier its C-terminal part (close to helix C).

Fig. 4 and Fig. 5 also show a delayed initiation of structural change in residues 4–7 of helix A, residues 27–30 of binding loop I and N-terminal part of helix B. Specifically, the residues near the top part of helix B (close to binding loop I) and in binding loop I, have very little structural change at the beginning of domain opening, with a sharp, cooperative transition near the end. [See the relatively slow color change (from red to blue) in this part of helix B and binding loop I in Fig. 5(a)-(d).] Although, the middle part of helix C (residues 50–52) has some limited structural change early in the transition, it remains quite immobile

after that. [See Fig. 4 and the early color change from red to blue in Fig. 5.]

Binding loops I and II. Because of the central importance of the interactions between the binding loops in the recently proposed two-step Ca^{2+}-binding mechanism, this EFβ-scaffold region is highlighted in Fig. 6. In the first step of this binding mechanism, the Ca^{2+} is immobilized by the structural rigidity in the plane of β-sheet and the ligands from N-terminal part of the binding loops. In the second step, the backbone torsional flexibility of the EFβ-scaffold enables repositioning of the C-terminal part of the binding loop together with the exiting helix (helix B in loop I and helix D in loop II).[13] Since the Ca^{2+} ions are not included in our model and we can not characterize backbone torsional flexibility of the EFβ-scaffold, our analysis is independent of that developed in Ref. 12,13. The closed to open conformational transition of each binding loop is quite different in Fig. 6. We predict that the structural changes in binding loop II occur before binding loop I upon domain opening (see the relatively slow color change from red to blue in binding loop I than loop II in Fig. 6). Since the flexibility of binding loop II is also greater, this suggests that during Ca^{2+}-binding process the loop II is more dominates the overall conformational change between the closed and open state. This agrees with results based on the all atom molecular dynamics simulations of nCaM discussed by Vigil et al..[20]

Fig. 6 also shows that the N-terminal ends of the loops have relatively an early transition compared to the C-terminal ends. Furthermore, the conformation change of the C-terminal end of binding loop I is more cooperative, presumably relying on the earlier structural change in binding loop II. Specifically, the closed state structure residue in position 9 (Thr28) of the loop I is very stable as shown in Fig. 7(a). This is due to a hydrogen bonding between Thr28 and Glu31. Fig. 7(a) also suggests that the structural change of Glu31 occurs before Thr28 upon domain opening, and proceeds through the transition much more gradually. Similar hydrogen bonding is also present between Asn64 and Glu67 in binding loop II. Nevertheless, compared to the corresponding residues in loop I, the structural change of these two residues is quite gradual [see Fig. 7(a)]. Nevertheless, Asn64 does seem to have a somewhat sharper transition than Glu67. Finally, residues Gly61 and Thr62 in binding loop II exhibit little structural change in Fig. 6 as the domain begins to open.

Methionine residues. The large hydrophobic binding surfaces that open in both domains of CaM are especially rich in Methionine residues, with four Methionines in each domain occupying nearly 46% of the total hydrophobic surface area.[3] These side chains as

well as other aliphatic residues, such as Valine, Isoleucine and Leucine, which make up the rest of the hydrophobic binding surface are highly dynamic in solution.[32] The flexibility of the residues composing hydrophobic binding surface for target peptides explains CaM's high degree of binding promiscuity. Here we consider the main-chain flexibility. The four Methionine residues in nCaM are situated in position 36, 51, 71 and 72. The closed to open structural transition of residues Met36 and Met71 are similar and relatively sharp compared to residue Met72 which is quite gradual as shown in Fig. 7(b). This suggests that residues Met36 and Met71 remains relatively buried in the beginning of the domain opening. Curiously, from Fig. 7(b) residue Met51 in the middle part of helix C at $\alpha_0 = 0.5$, shows sudden increase in $\Delta\overline{\rho_i}$ during closed to open conformational change.

Conformational Transition Rate and Order Parameter

The one dimensional free energy profile parameterized by the interpolation parameter α_0 is shown in Fig. 8. The minimum corresponding to the open state is very shallow and unstable compared to the closed state. Combined molecular dynamics simulations and small angle X-ray scattering studies on apo nCaM and Ca^{2+}-bound nCaM by Vigil et al.[20] have also shown that in aqueous solution the closed state dominates the population. The equilibrium populations for the closed and open state from our model are found to be 94% and 6% respectively. For comparison, the NMR measurement of apo cCaM indicate a minor population of 5–10%.[9] These results suggest that on average, the residues in the hydrophobic surface of CaM are well protected from solvent.

The maximum of the free energy occurs quite close to the open state at $\alpha_0 \sim 0.2$, though the barrier is very broad in terms of this reaction coordinate. We also consider the free energy of the global structural parameter $\Delta Q = Q_1 - Q_2 = \sum \Delta\overline{\rho_i}/N$ where $\Delta\overline{\rho_i}$ is given in in Eq. 11. Fig. 8 shows that ΔQ is also a reasonable reaction coordinate for the transition. The barrier broadens somewhat, with the maximum free energy occurring around $\Delta Q = -0.25$. In terms of the global structure, this roughly corresponds to 60%–75% of nCaM being similar to open state configuration in the transition state ensemble.

Even though the open state minimum is not well isolated, we estimate the conformational transition rate from closed to open using the Arrhenius form, $k = k_0 e^{-\Delta F^\dagger/k_B T}$ where $\Delta F^\dagger$ is the free energy difference between the closed conformation and transition-state ensemble.

Assuming the prefactor $k_0 = 1\mu s^{-1}$ gives the estimate $k = 40,000 s^{-1}$. This value is in reasonable agreement with the transition rate estimate of $k = 20,000 s^{-1}$ based on NMR exchange rate data of cCaM.[9]

DISCUSSION

The primary motivation for the work presented in this paper is to understand protein functions that involve large scale (main-chain) dynamics and flexibility. Proteins with relatively large conformational freedom include those in which folding and binding are coupled.[33], as well as hinge bending motions[34] or proteins with high plasticity such as ion binding sites,[35] and proteins with allosteric transitions.[36] While not nearly as developed as the Energy Landscape Theory of protein folding,[37] a general thermodynamic framework for the Energy Landscape Theory of protein-protein binding,[38,39] large conformational transitions,[34] and the coupling between folding and binding[40] is beginning to emerge. Aside from some noted exceptions,[25,41,42,43,44,45] relatively little theoretical work has focused on detailed analysis of transition mechanisms of flexible proteins in terms of specific ensembles of kinetic pathways. The dynamics of conformational transitions between well-defined conformational basins are generally controlled by relatively low probability partially ordered ensembles. The main challenge is to describe the transition state ensembles at the residue level giving a site-specific description of the transition mechanism.

Modern NMR relaxation experiments have provided a wealth of data about internal dynamics and conformational sub-states quantitatively on fast (nanosecond) and slow (micro- to millisecond) timescales.[10] Such studies are very useful in identifying residues with high flexibility upon target binding, not only through movements of surface loops and side chains, but also by global motions of the core structure.[46] These experiments, however, provide only a few local structural changes and have not been able to capture the molecular details necessary to fully understand the mechanism of conformational transitions. Whereas atomistic simulations can potentially bridge the gap on time scale up to microsecond, this timescale falls orders of magnitude short for slow protein dynamics (millisecond to second). Also, the use of atomistic approaches becomes computationally inefficient with the increased size of a system.

To overcome the problems associated with all-atom simulations, many studies has demon-

strated the use of coarse-grained protein models with simplified representations, such as, only α-carbons as point masses and simplified energy functions.[47] Such models require much less computational cost making them practical to describe the conformational transitions of even large proteins.[28] Analyzing the fluctuations about a single minimum has been surprisingly successful in identifying relevant cooperative motions in a wide range of proteins. The commonly used Tirion potential[48,49] (which can be viewed as a harmonic Go-model) gives a simple one parameter model in which the relevant motions for the transition is identified as one of many low frequency normal modes.[50] While this approach can provide considerable insight, it offers a limited description of the transition because it is based only on the fluctuations about one structure. The Tirion potential has recently been extended to include two conformations in which the contact map defining the potential and normal modes is updated as the protein is moved along a known reaction coordinate.[42,51] Local unfolding and flexibility is accommodated by relieving regions of high stress, "cracking", which modifies the contact map. Coarse-grained simulations in which the potential interpolates between two folded-state biased contact maps have also been introduced recently.[21,24,25,26] For example, in the plastic network model of Margakis and Karplus[26] the individual basins are approximated by the Tirion potential and are then smoothly connected by a secular equation formulation. A similar interpolation was considered by Okazaki et al.[24] Alternatively, Best et al. developed a two-state approximation[25] analogous to Eq. 6. These advances are similar in spirt to our approach, albeit with distinct approximations for the basic description of partially ordered ensembles.

CONCLUSION

In this paper, we study the intrinsic flexibility and structural change in the N-terminal domain of CaM (nCaM) during open to close transition. The predicted transition route from our model gives a detailed picture of the interplay between structural transition, conformational flexibility and function of N-terminal calmodulin (nCaM) domain. The results from our model are largely consistent with the important role that the immobile EFβ-scaffold region plays in the transition mechanism. Dissection of the transition route of this region further suggests that it is the early structural change of loop II that drives the cooperative completion of the interactions between the loops in the open structure.

The strong qualitative agreement with available experimental measurements of flexibility is an encouraging validation of the model. Recently, the folding dynamics of zinc-metallated protein (azurin) was studied using a similar variational model and compared with experiments for the detail coordination reaction coupled with the entatic state.[52] A similar future study of detail coordination reaction for the complete description of conformational change stabilized by ion binding in CaM seems very promising. Ultimately, we wish to extend this model to investigate the binding mechanism and kinetic paths of several peptides to Ca^{2+}-loaded CaM. Since large conformational changes coupled to binding depends fundamentally on the fluctuations of partially folded conformations,[33] this polymer based variational formalism can accommodate coupled folding and binding very naturally.

Acknowledgments

We thank Zenon Grabarek for helpful suggestions and critically reading the manuscript. This work was supported in part by grant awarded by the Ohio Board of Regents Research Challenge program.

REFERENCES

[1] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991).

[2] M. Gerstein, A. M. Lesk, and C. Chothia, Biochemistry **33**, 6739 (1994).

[3] M. R. Nelson and W. J. Chazin, Protein Sci. **7**, 270 (1998).

[4] R. D. Brokx, M. M. Lopez, H. J. Vogel, and G. Makhatadze, J. of Biol. Chem. **276**, 14083 (2001).

[5] J. J. Chou, S. Li, C. B. Klee, and A. Bax, Nat. Struct. Biol. **8**, 990 (01).

[6] T. N. Tsalkova and P. L. Privalov, J. Mol. Biol. **181**, 533 (1985).

[7] S. Linse, A. Helmersson, and S. Forsen, J. Biol. Chem. **266**, 8050 (1991).

[8] J. L. Baber, A. Szabo, and N. Tjandra, J. Am. Chem. Soc. **123**, 3953 (2001).

[9] A. Malmendal, J. Evanas, S. Forsen, and M. Akke, J. Mol. Biol. **293**, 883 (1999).

[10] R. Ishima and D. A. Torchia, Nat. Struct. Biol. **7**, 740 (2000).

[11] J. Evenas, S. Forsen, A. Malmendal, and M. Akke, J. Mol. Biol. **289**, 603 (1999).

[12] Z. Grabarek, J. Mol. Biol. **346**, 1351 (2005).

[13] Z. Grabarek, J. Mol. Biol. **359**, 509 (2006).

[14] H. Kuboniwa, N. Tjandra, S. Grzesiek, H. Ren, C. B. Klee, and A. Bax, Nat. Struct. Biol. **2**, 768 (1995).

[15] M. Zhang, T. Tanaka, and M. Ikura, Nat. Struct. Biol. **2**, 758 (1995).

[16] J. J. Portman, S. Takada, and P. G. Wolynes, Phys. Rev. Lett. **81**, 5237 (1998).

[17] J. J. Portman, S. Takada, and P. G. Wolynes, J. Chem. Phys. **114**, 5069 (2001).

[18] J. J. Portman, S. Takada, and P. G. Wolynes, J. Chem. Phys. **114**, 5082 (2001).

[19] W. Wriggers, E. Mehler, F. Pitici, H. Weinstein, and K. Schulten, Biophys. J. **74**, 1622 (1998).

[20] D. Vigil, S. C. Gallagher, J. Trewhella, and A. E. Garcia, Biophys. J. **80**, 2082 (2001).

[21] D. M. Zuckerman, J. Phys. Chem. B **108**, 5127 (2004).

[22] Y.-G. Chen and G. Hummer, J. Am. Chem. Soc. **129**, 2414 (2007).

[23] M. Bixon and R. Zwanzig, J. Chem. Phys. **68**, 1896 (1978).

[24] K. Okazaki, N. Koga, S. Takada, J. N. Onuchic, and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **103**, 11844 (2006).

[25] R. B. Best, Y. G. Chen, and G. Hummer, Structure **13**, 1755 (2005).

[26] P. Maragakis and M. Karplus, J. Mol. Biol. **352**, 807 (2005).

[27] S. Miyazawa and R. L. Jernigan, J. Mol. Biol. **256**, 623 (1996).

[28] M. K. Kim, R. L. Jernigan, and G. S. Chirikjian, Biophys. J. **83**, 1620 (2002).

[29] R. B. Russell and G. J. Barton, Proteins Struct. Funct. Genet. **14**, 309 (1992).

[30] V. A. Likic, E. E. Strehler, and P. R. Gooley, Protein Sci. **12**, 2215 (2003).

[31] N. C. J. Strynadka and M. N. James, Annu. Rev. Biochem. **58**, 951 (1989).

[32] K. Siivari, M. Zhang, I. Arthur G. Palmer, and H. J. Vogel, FEBS Lett. **366**, 104 (1995).

[33] B. A. Shoemaker, J. J. Portman, and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **97**, 8868 (2000).

[34] N. Sinha, S. Kumar, and R. Nussinov, Structure **9**, 1165 (2001).

[35] A. Lewit-Bentley and S. Rety, Curr. Opin. Struct. Biol. **10**, 637 (2000).

[36] P. Lundstrom, F. A. A. Mulder, and M. Akke, Proc. Natl. Acad. Sci. USA **102**, 16984 (2005).

[37] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins Struct. Funct. Genet. **21**, 167 (1995).

[38] G. M. Verkhivker, D. Bouzida, D. K. Gehlhaar, P. A. Rejto, S. T. Freer, and P. W. Rose, Curr. Opin. Struct. Biol. **12**, 197 (2002).

[39] J. Wang, K. Zhang, H. Y. Lu, and E. K. Wang, Phys. Rev. Lett. **96**, 168101 (2006).

[40] G. A. Papoian and P. Wolynes, Biopolymers **63**, 333 (2003).

[41] G. M. Verkhivker, D. Bouzida, D. K. Gellhaar, P. A. Tejto, S. T. Freer, and P. W. Rose, Proc. Natl. Acad. Sci. USA **100**, 5148 (2003).

[42] O. Miyashita, J. N. Onuchic, and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **100**, 12570 (2003).

[43] Y. Levy, P. G. Wolynes, and J. N. Onuchic, Proc. Natl. Acad. Sci. USA **101**, 511 (2004).

[44] S. Yang, S. S. Cho, Y. Levy, M. S. Cheung, H. Levine, P. G. Wolynes, and J. N. Onuchic, Proc. Natl. Acad. Sci. USA **101**, 13786 (2004).

[45] Y. Levy, S. S. Cho, J. N. Onuchic, and P. G. Wolynes, J. Mol. Biol. **346**, 1121 (2005).

[46] Y. J. Huang and G. T. Montelione, Nature **438**, 36 (2005).

[47] V. Tozzini, Curr. Opin. Struct. Biol. **15**, 144 (2005).

[48] M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).

[49] I. Bahar, A. R. Atilgan, M. C. Demirel, and B. Erman, Phys. Rev. Lett. **80**, 2733 (1998).

[50] F. Tama and C. L. Brooks, J. Mol. Biol. **318**, 733 (2002).

[51] O. Miyashita, P. G. Wolynes, and J. N. Onuchic, J. Phys. Chem. B **109**, 1959 (2004).

[52] C. Zong, C. J. Wilson, T. Shen, P. Wittung-Stafshede, S. L. Mayo, and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **104**, 3159 (2007).

[53] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).

Figure Legends

Figure 1.

The N-terminal domain of calmodulin (nCaM). (a) The Ca^{2+}-free (apo, closed) structure, PDB code 1cfd. (b) The Ca^{2+}-bound (holo, open) structure, PDB code 1cll. (c) The secondary structure of nCaM is shown with one letter amino acid sequence code for residues 4-75. The secondary structure of nCaM is as follows: helix A (5–19), Ca^{2+}-binding loop I (20–31), helix B (29–37), B/C helix-linker (38–44), helix C (45–55), Ca^{2+}-binding loop II (56–67), helix D (65–75). Note that, the last three residues of the binding loops I and II are also part of the exiting helices B and D. There are short β-sheet structures in binding loop I (residues 26–28) and loop II (residues 62–64). This, and other three-dimensional illustrations were made using Visual Molecular Dynamics (VMD).[53]

Figure 2.

Fluctuations $B_i = \langle \delta \mathbf{r}_i^2 \rangle_0 = G_{ii} a^2$ vs sequence index of nCaM for selected values of the interpolation parameter α_0 in the conformational transition route between open and closed. Here $a = 3.8$Åis the distance between successive monomers. Different α_0 are denoted by, red ($\alpha_0 = 0$) open; green ($\alpha_0 = 0.2$); blue ($\alpha_0 = 0.4$); pink ($\alpha_0 = 0.6$); orange ($\alpha_0 = 0.8$) and black ($\alpha_0 = 1$) closed. The secondary structure is indicated below the plot.

Figure 3.

Change in fluctuations B_i in nCaM domain during the closed to open conformational transition. The 3D structure in (a) corresponds to the interpolation parameter, $\alpha_0 = 1$ (closed state); (b) corresponds to $\alpha_0 = 0.4$ (intermediate state) and (c) corresponds to $\alpha_0 = 0$ (open state). Red corresponds to low fluctuations and blue corresponds to high. Here, a is the distance between successive monomers.

Figure 4.

Difference between the normalized native density $\Delta\overline{\rho_i}$ (a measure of structural similarity) of each residue for different α_0. The change in color from red to blue is showing the closed $\rightarrow$ open conformational transition of nCaM. This is normalized to be -1 at the open state minimum ($\alpha_0 = 0$; blue) and 1 at the closed state minimum ($\alpha_0 = 1$; red). Below the secondary structure of nCaM is shown. Here, α^N in Eq. 7 and Eq. 8 is 0.5.

Figure 5.

Closed to open conformational transition in nCaM with different interpolation parameter α_0. The 3D structure in (a) corresponds to the interpolation parameter, $\alpha_0 = 0.8$; (b) corresponds to $\alpha_0 = 0.6$; (c) corresponds to $\alpha_0 = 0.4$ and (d) corresponds to $\alpha_0 = 0.2$. The change in color from red to blue corresponds to different values of normalized native density $\Delta\overline{\rho_i}$ (a measure of structural similarity) of each residue for different α_0. Red corresponds to $\Delta\overline{\rho_i} = 1$ (closed conformation) and blue (open conformation) corresponds to $\Delta\overline{\rho_i} = -1$.

Figure 6.

Comparison of structural change in binding loops I (in bottom) and II (in top) in terms of the order parameter $\Delta\overline{\rho_i}$. The 3D structures in (a)-(i) corresponds to the interpolation parameter, $\alpha_0 = 0.9$ -0.1 during the closed to open transition. The change in color from red to blue corresponds to different values of $\Delta\overline{\rho_i}$ (a measure of structural similarity) of each residue. Red corresponds to $\Delta\overline{\rho_i} = 1$ (closed conformation) and blue (open conformation) corresponds to $\Delta\overline{\rho_i} = -1$.

Figure 7.

Dynamical behavior of residues during conformational transition of nCaM. The normalized native density difference $\Delta\overline{\rho_i}$ vs α_0 are shown for four different group of residues. Structural transition of (a) residues in position 9 (Thr28 and Asn64) and position 12 (Glu31 and Glu67) of the two binding loops; (b) four hydrophobic Methionine residues in positions 36, 51, 71 and 72.

Figure 8.

Free energy along the transition route. In the lower curve the abscissa is the interpolation parameter α_0. In the upper curve the abscissa is the global structural order parameter ΔQ. The entropy across the transition is relatively constant, so that the free energy barrier is largely energetic.

Figures

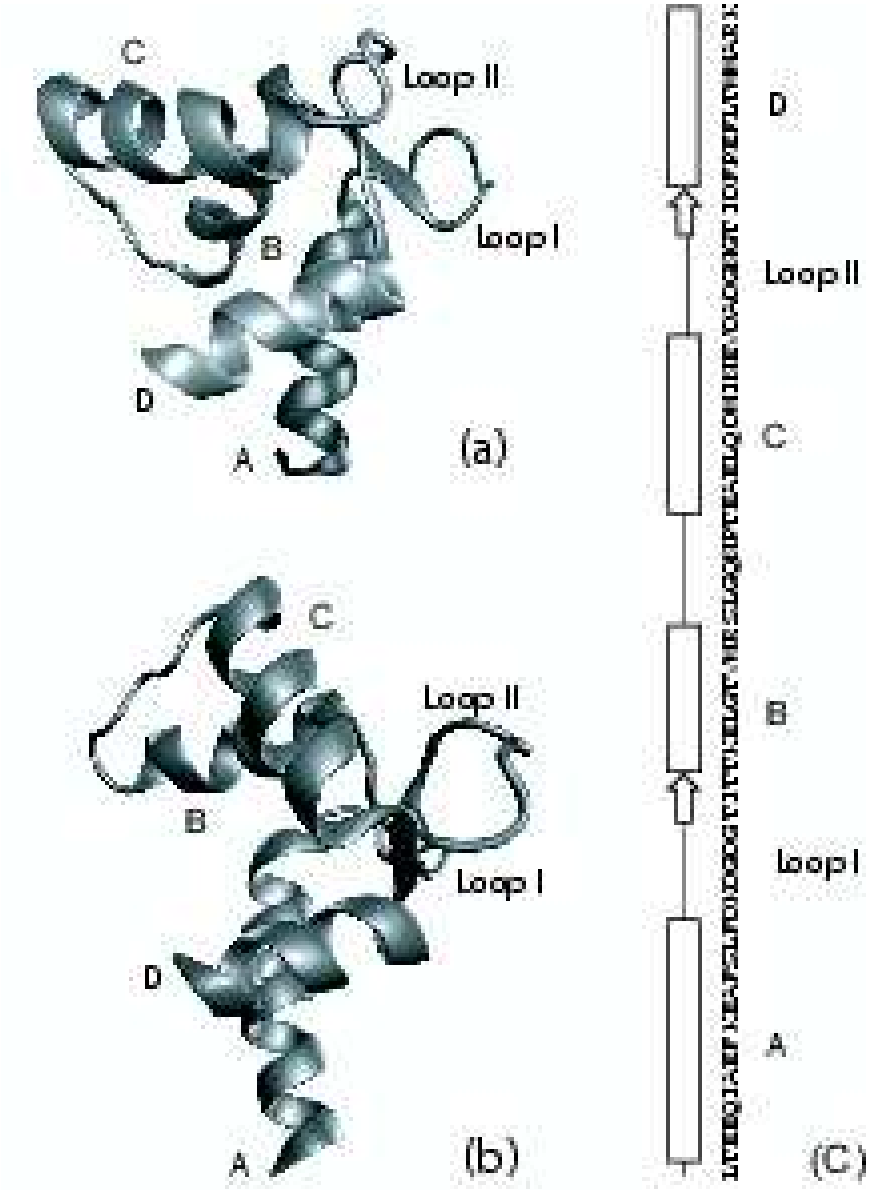

FIG. 1:

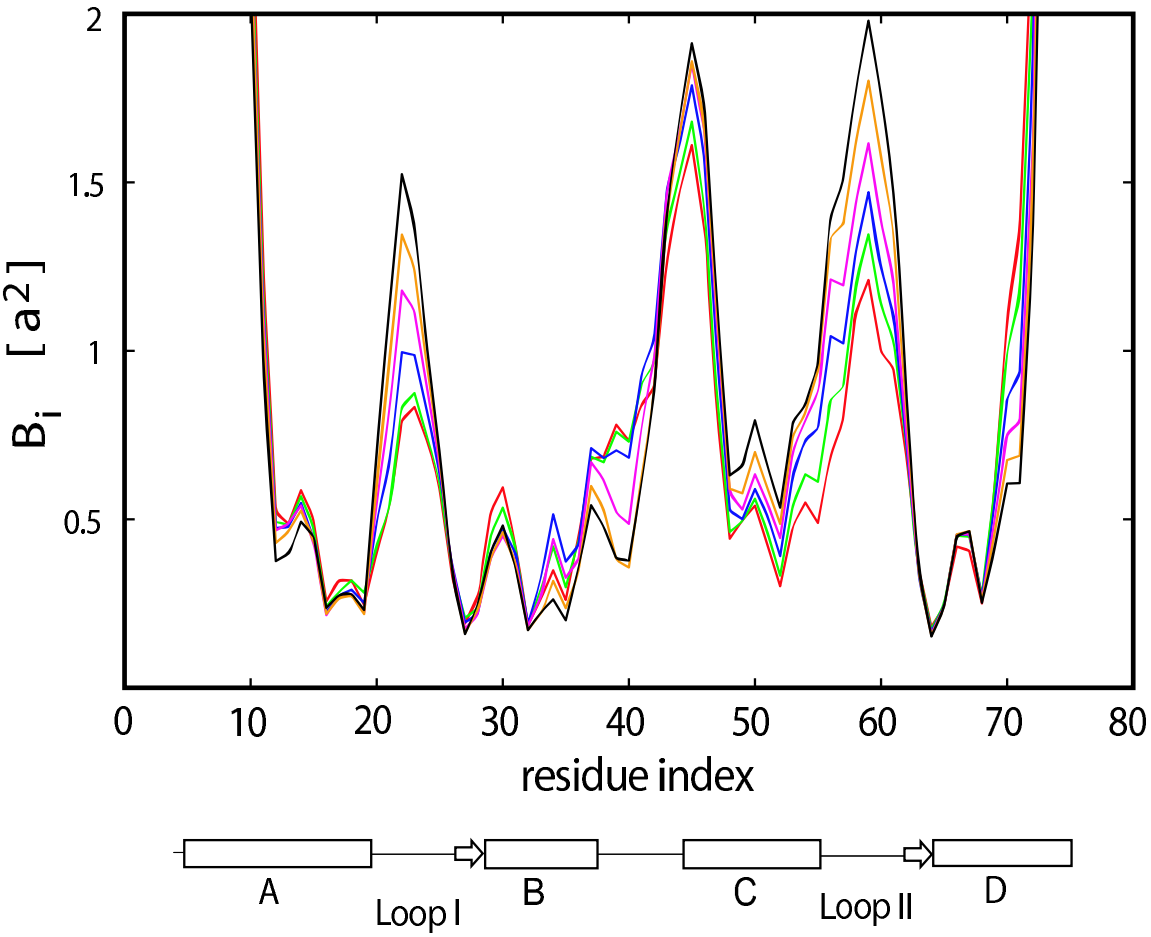
2
1.5
1
0.5
B_i [a^2]
0
10
20
30
40
50
60
70
80
residue index
A
Loop I
B
C
Loop II
D

FIG. 2:

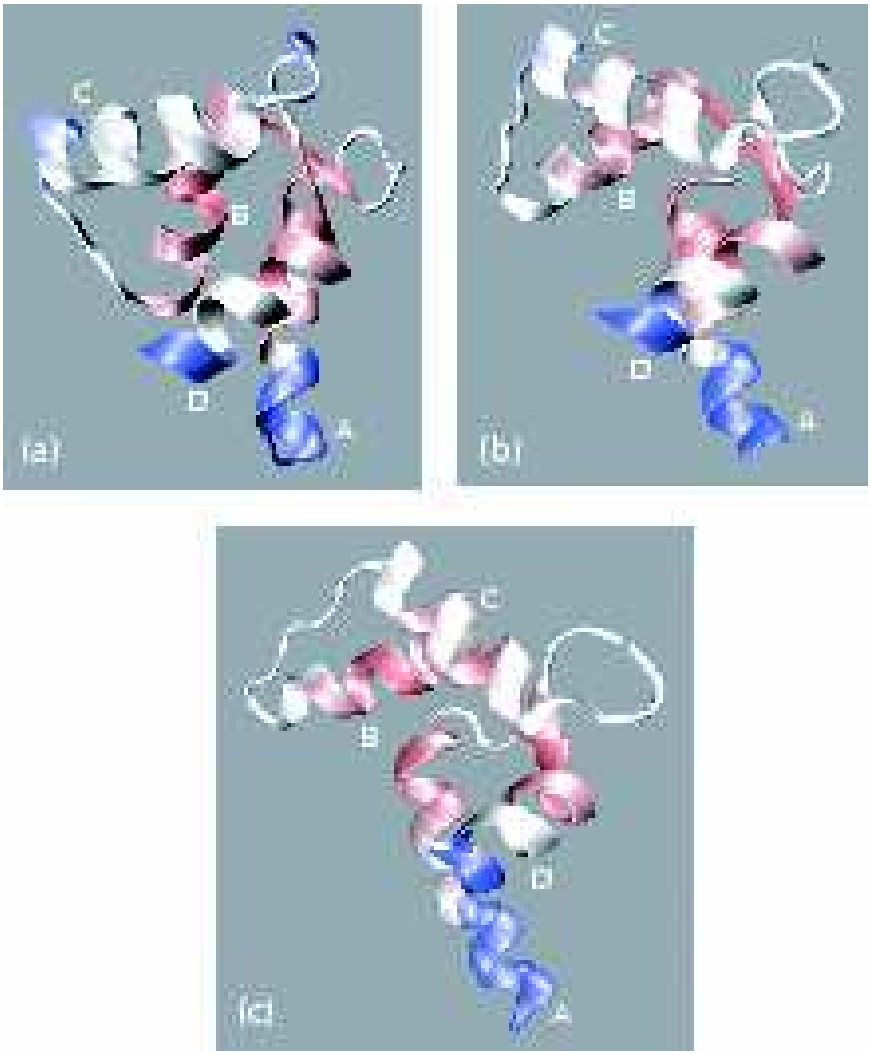

FIG. 3:

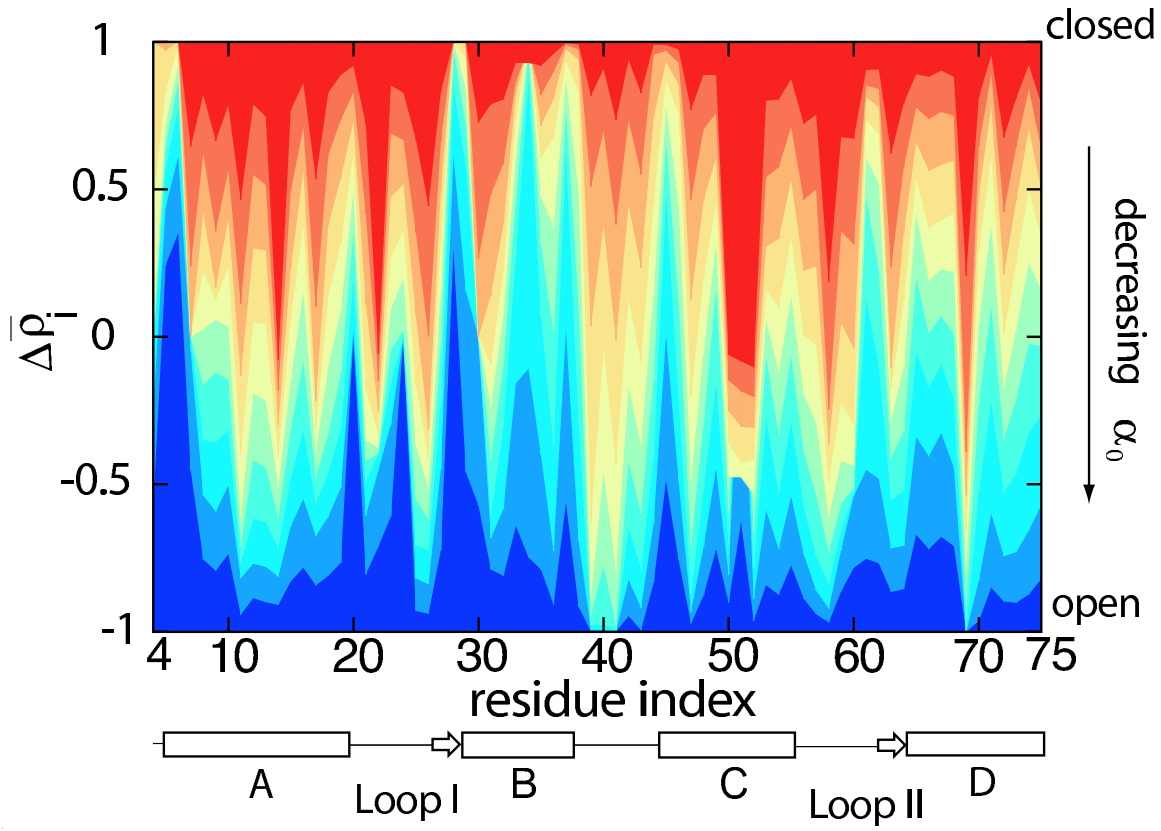
1
0.5
0
-0.5
-1
$\Delta\bar{\rho}_i$
closed
decreasing α_0
open
4
10
20
30
40
50
60
70
75
residue index
A
Loop I
B
C
Loop II
D

FIG. 4:

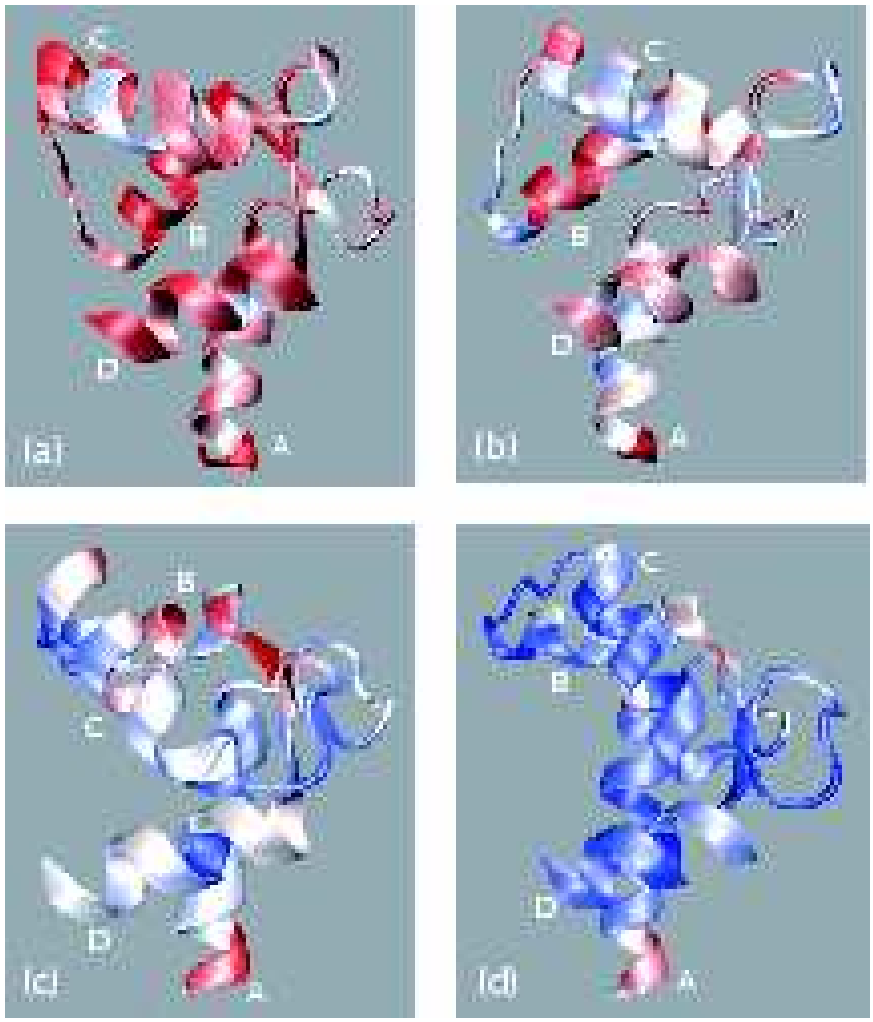

FIG. 5:

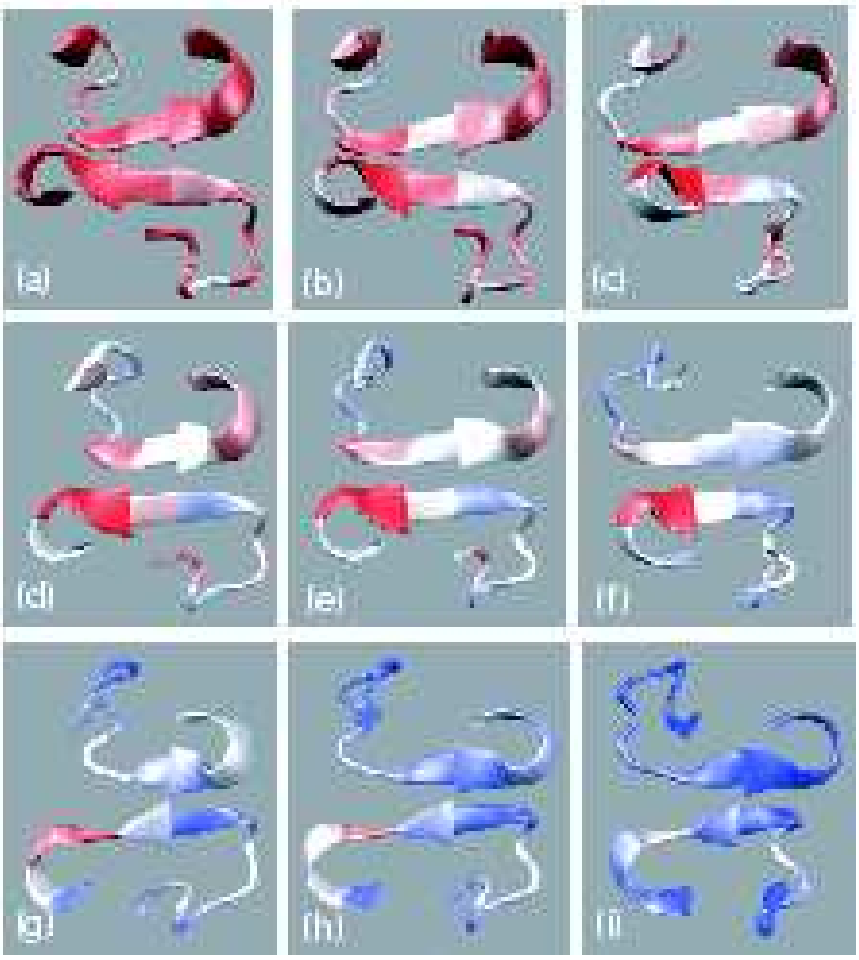

FIG. 6:

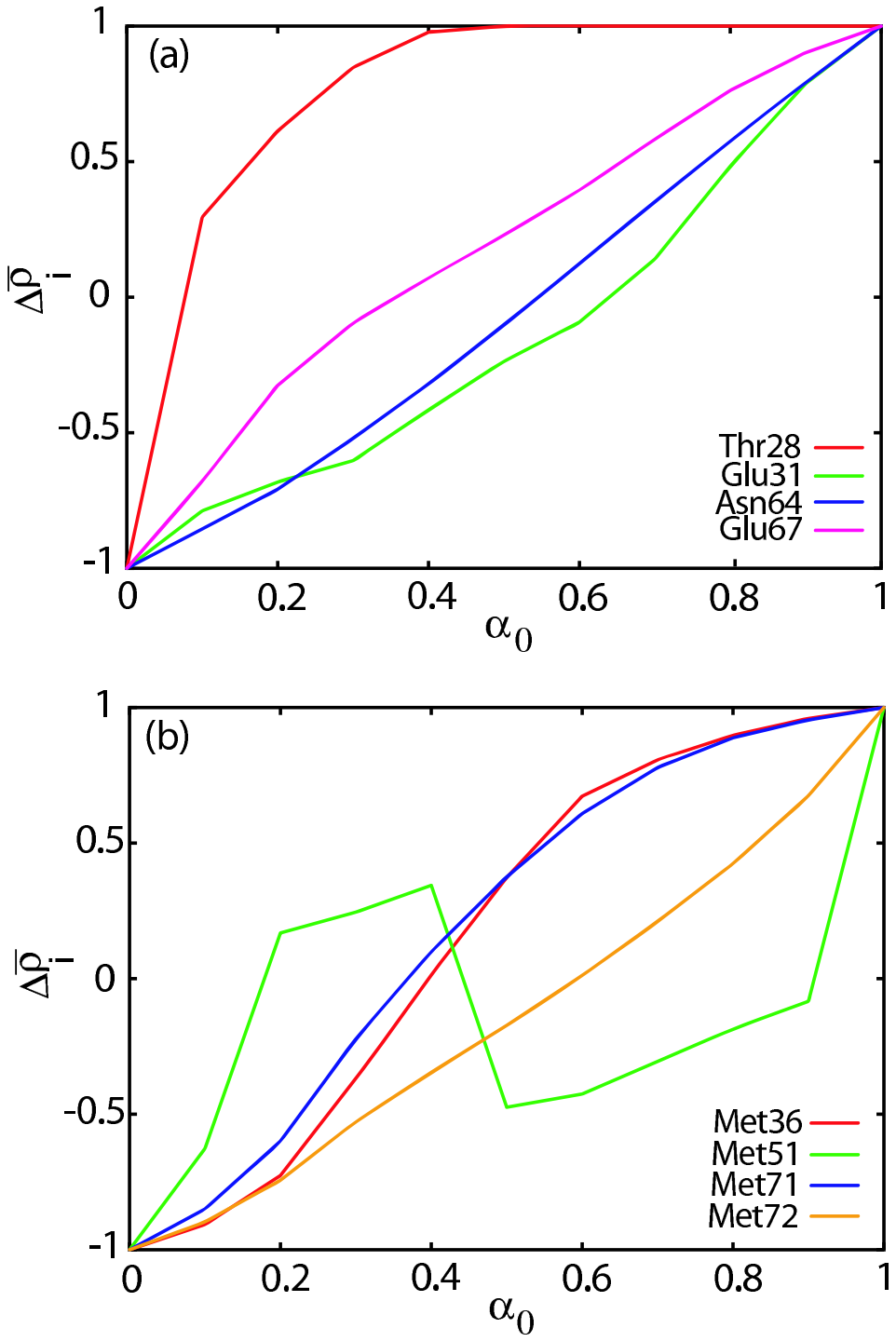
(a)
1
0.5
0
-0.5
-1
$\Delta\bar{\rho}_i$
0
0.2
0.4
0.6
0.8
1
α_0
Thr28
Glu31
Asn64
Glu67
(b)
1
0.5
0
-0.5
-1
$\Delta\bar{\rho}_i$
0
0.2
0.4
0.6
0.8
1
α_0
Met36
Met51
Met71
Met72

FIG. 7:

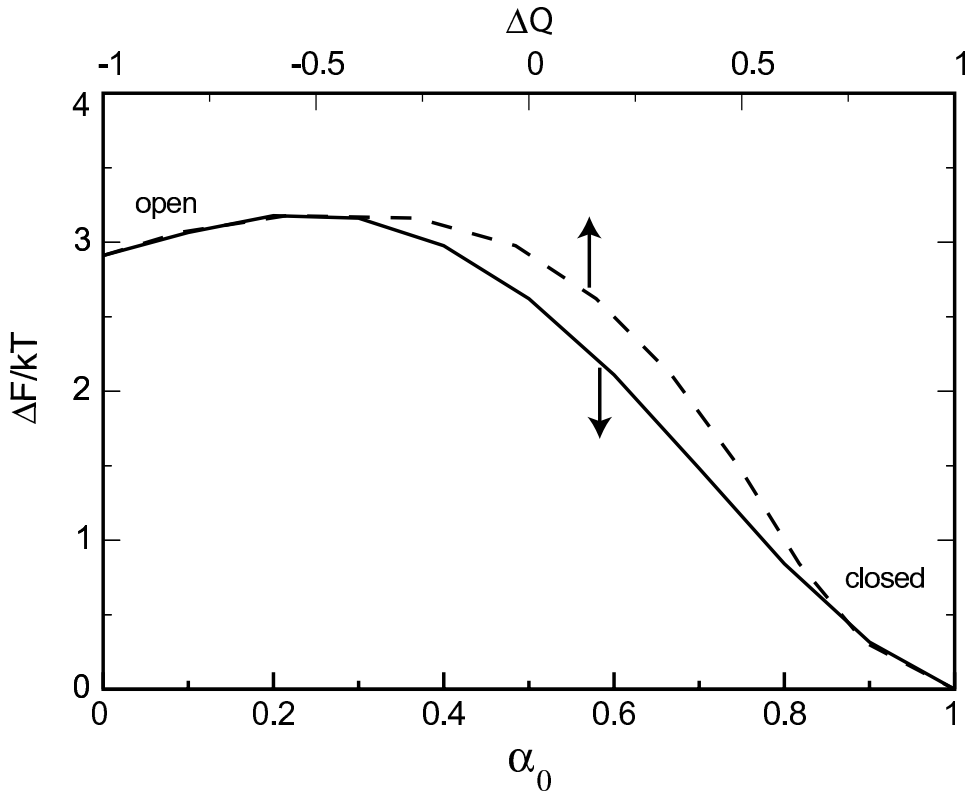

FIG. 8: