Molecular Recognition Mechanisms of Calmodulin
Examined by Perturbation-Response Scanning

A. Ozlem Aykut[1], Ali Rana Atilgan[1], Canan Atilgan[1]

[1] Faculty of Engineering and Natural Sciences, Sabanci University, 34956 Istanbul, Turkey

ABSTRACT

We analyze the apo and holo calmodulin (CaM) structures by sequentially inserting a perturbation on every residue of the protein, and monitoring the linear response. Residue cross-correlation matrices obtained from 20 ns long molecular dynamics simulation of the apo-form are used as the kernel in the linear response. We determine two residues whose perturbation equivalently yields the experimentally determined displacement profiles of CaM, relevant to the binding of the trifluoperazine (TFP) ligand. They reside on structurally equivalent positions on the N- and C-terminus lobes of CaM, and are not in direct contact with the binding region. The direction of the perturbation that must be inserted on these residues is an important factor in recovering the conformational change, implying that highly selective binding must occur near these sites to invoke the necessary conformational change.

INTRODUCTION

Calmodulin (CaM) has a pivotal role as an intracellular Ca^{2+} receptor that is involved in calcium signaling pathways in eukaryotic cells [1]. CaM can bind to a variety of proteins or small organic compounds, and can mediate different physiological processes by activating various enzymes [2,3]. Binding of Ca^{2+} and proteins or small organic molecules to CaM induces large conformational changes that are distinct to each interacting partner [1,3,4]. It has recently been shown that humans possess three CaM genes that are differentially regulated and which encode identical proteins [3]. Diseases related to unregulated cell growth, such as cancer, have been shown to have elevated levels of Ca^{2+} loaded CaM (Ca^{2+}-CaM) [3]. To design drugs that inhibit Ca^{2+}-CaM formation, the molecular binding mechanism must be thoroughly understood.

In this study, we use the perturbation-response scanning (PRS) tool that is based on systematically exerting directed forces on the residues of the protein [5] and recording the response as changes in residue fluctuation profiles. We find that by perturbing single residues located at a variety of positions in the apo-form, it is possible to induce different conformational changes relevant to the binding of ligands. Depending on the topology of the particular binding site, the unbound form may be manipulated towards the bound form by directly perturbing select residues on the distant helices in preferred directions. Our findings also give information on how the flexible linker region acts as a transducer of binding information to distant parts of the protein. This study is a first step towards revealing the essence of ligand recognition mechanisms by which CaM controls a wide variety of Ca^{2+} signaling processes.

THEORY AND METHODS

Perturbation response scanning (PRS) method is based on inserting forces as perturbations on a given residue, and recording the displacements of all the residues as the response. We perceive the protein as a residue network of N nodes centered on the C_α atoms.

Any given pair of nodes are assumed to interact via a harmonic potential, if they are within a cut-off distance r_c of each other. One gets response of the residue network given to perturbations by:

$$\mathbf{H}^{-1}\Delta\mathbf{F} = \Delta\mathbf{R} \qquad (1)$$

where the $\Delta\mathbf{F}$ vector contains the components of the externally inserted force vectors on the selected residues. Details leading to the above equation can be found in refs. 5 and 6. Elements of the inverse of the Hessian, $\mathbf{G} = \mathbf{H}^{-1}$, may be used to predict the auto- and cross-correlations of residues. $\mathbf{G}$ is a 3N×3N matrix and may be viewed as an $N \times N$ matrix whose ijth element is a 3×3 matrix of correlations between the x-, y-, and z-components of the fluctuations $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_j$ of residues i and j. The cross-correlations between residue pairs are obtained from the trace of its components; $\langle\Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j\rangle = tr(\mathbf{G}^{ij})/N$. This relation has been shown to reproduce the cross-correlations obtained from molecular dynamics (MD) simulations and molecular mechanics [10, 11]. Therefore, one may also use cross-correlation profiles obtained from MD as a kernel in eq.1.

Our detailed PRS analysis is based on a systematic application of eq. 1. We apply a force, mimicking small ligand binding, on the C_α atom of each residue by forming the $\Delta\mathbf{F}$ vector in such a way that all the entries, except those corresponding to the residue being perturbed, are equal to zero. For a selected residue i, the random force q_i is (q_x, q_y, q_z) so that the external force vector is constructed as, $(\Delta\mathbf{F})^{\mathrm{T}} = \{0\ 0\ 0\ ...\ q_x\ q_y\ q_z\ ...\ 0\ 0\ 0\}_{1\times 3N}$. The direction of the force is chosen randomly, attributing no bias due to the specific contact topology or the solvent exposed nature of the residue being perturbed. Force directions are uniformly distributed within a sphere enveloping the residue; therefore, the forcing may well be termed isotropic. The details of the force application procedure is given in reference 5. We then compute the resulting $(\Delta\mathbf{R})^i$ vector of the protein through eq. 1. We scan the protein by consecutively perturbing each residue, and recording the expected displacements as a result of the linear response of the protein [5]. In all calculations based on eq. 1, we report the averages over ten independent runs where the forces are applied randomly [5, 8]. Thus, the predictions of the average displacement of each residue, j, as a response of the system to inserted forces on residue i, $(\Delta R_j)^i$ are compared with the experimental conformational changes between the original and the target PDB structure, i.e. the apo and the holo forms. The goodness of the prediction is quantified as the Pearson correlation (PC) and overlap coefficient (OL) for each perturbed residue. The former is a measure of the relative displacement of each residue in response to the perturbation, while the latter is the dot product of the two vectors, as a measure of the similarity of both the direction and the magnitude in the predicted conformational change.

PRS technique produces both the magnitude and the directionality of the response. However, averaging over several independent runs focuses on capturing the relative magnitudes of the displacements and disregards the directionality information. Note that if the collection of forces applied on a specific residue is independent and large in number, they will appear in a spherically symmetric set of directions. The responses, however, need not be recovered in the same manner; e.g. they may be distributed along a line or in a plane so that the net response is still zero [5, 8]. Deviations from such a spherically symmetric distribution of responses hint at the roles of certain residues in the remote control of the ligand. The responses given to perturbation of the residues yielding high OL are visualized as vector plots on the protein.

In the present study, we study the conformational change between the calcium bound CaM in its free and trifluoperazine (TFP) bound forms; the protein data bank codes are 3cln and

1lin, respectively (fig. 1a,b). CaM is a 148 residue protein and has two structurally similar domains. Each domain consists of two helix-loop-helix Ca^{2+}-binding regions referred to as EF-hand structure. There is evidence that the long alpha-helix joining the two domains is present only in the crystal structure, whereas, in solution, this region is extremely flexible [3]. Upon binding to the ligand, the domains move relative to each other and the flexible linker region is manipulated accordingly. Unless otherwise specified, the final structures are superimposed on all residues in the initial structure before the displacements are computed in the PRS method. Both the apo and holo forms contain four Ca^{2+} ions, two bound to each lobe. Four TFP molecules are bound to the holo form, lying between the two domains [3,12].
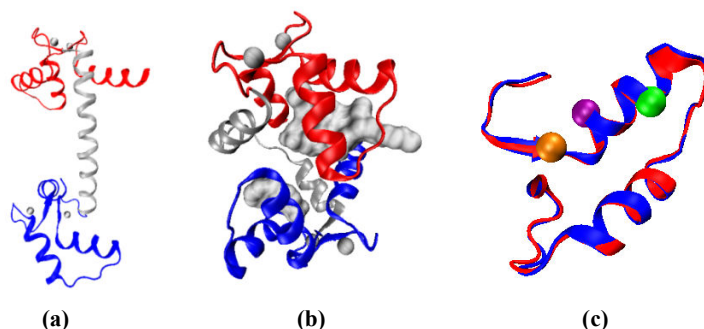

| (a) | (b) | (c) |

Figure 1. X-ray crystal structure of the proteins utilized in this work: (a) 3cln (b) 1lin. N-terminal domain is in blue and C- terminal domain is in red. The Ca^{2+} ions are shown as gray spheres, and the bound TFP molecules are shown in gray space filling representations in (b). Superposition of N- (residues 22-63) and C-terminal (residues 93-136) lobes is shown in (c). Marked residues on equivalently positioned helices are 27 (orange), 31 (purple), 109 (green).

We use a cut-off distance of $r_c = 10.5$ Å on the C_α atoms of the protein and the four Ca^{2+} ions to generate the residue network of the apo-form; the inverse Hessian from this approach is abbreviated as $\mathbf{G}_{net}$. Alternatively, to provide a basis for comparison with results from an all-atom approach, we have performed MD simulations on the apo form in water. The systems are prepared using the VMD 1.8.7 program with solvate plug-in version 1.2 [13]. The NAMD package is used to model the dynamics of the protein – water systems [14]. TIP3P water model is used in the simulations. The protein is soaked in a solvent box so that the resulting system has dimensions $55.8 \times 73.5 \times 52.2$ Å^3. CharmM27 force field is used [15]. Long range electrostatic interactions are calculated using particle mesh Ewald method [16]. The cutoff distance for non-bonded interactions is set to 12 Å with a switching function cutoff of 10 Å. During the simulations, periodic boundary conditions are used and the time step size is 2 fs. Temperature control is carried out by Langevin dynamics with a dampening coefficient of 5/ps and pressure control is attained by a Langevin piston. Volumetric fluctuations are preset to be isotropic. System is subjected to energy minimization followed by the MD run in the NPT ensemble at 1 atm and 310 K for 20 ns. Averages are taken over the 1000 recorded snapshots and the $3N \times 3N$ residue cross correlation matrix is calculated via $\mathbf{G} = \Delta\mathbf{R}\,\Delta\mathbf{R}^T$. To evaluate the effect of trajectory length on the results, the first 10 ns of the trajectory and all of the 20 ns trajectory data are used to calculate separate correlation matrices (abbreviated as $\mathbf{G}_{10}$ and $\mathbf{G}_{20}$, respectively),

RESULTS AND DISCUSSION

The overall RMSD between the apo and holo forms is 14.9 Å. In contrast, the superimposition of only the N- or the C terminal yields RMSD of 0.63 Å (fig. 3c) and 0.70 Å, respectively for apo- and holo-forms, indicating that the local arrangements in the two lobes are nearly the same. This hints that the conformational change in going from apo to holo form

involves global motions instead of local rearrangements. Based on the linear response theory, we apply perturbations on each residue along the chain to diagnose the response of the protein. For every $\Delta\mathbf{R}$ vector obtained from eq. 1, we calculate PC and OL between this and the experimental $\Delta\mathbf{R}$ vector.

Figure 3 displays the PC and OL obtained from PRS using different Hessians as kernels (i) the network construction using 10.5 Å cutoff distance, $\mathbf{G}_{net}$ (ii) cross-correlations from 10 ns long trajectory, $\mathbf{G}_{10}$ (iii) cross-correlations from 20 ns long trajectory, $\mathbf{G}_{20}$. Each point on these figures is the result of the comparison of the displacements of the 148 residues in response to a perturbation applied at the selected residue, averaged over 10 independent directions. Note that the directional preference of the response (quantified by OL) is a more reliable measure for deciding whether the conformational change from apo to holo form may be achieved via PRS.
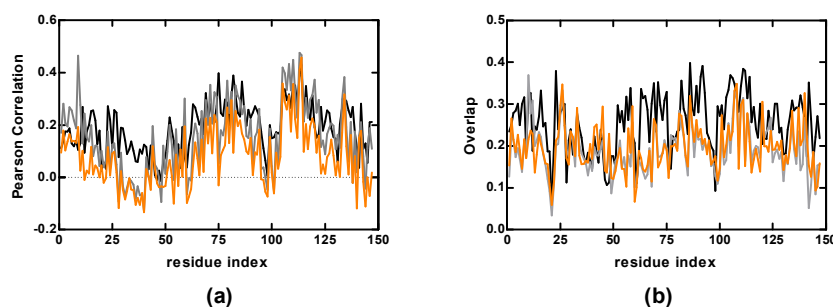

Figure 2. Magnitude (PC) and directional (OL) nature of the response. (Black: from $\mathbf{G}_{net}$, gray: from $\mathbf{G}_{10}$, orange: from $\mathbf{G}_{20}$)

All three kernels lead to similar results: Both the magnitude of the average response to perturbations (Fig. 2a), and the vectorial response (Fig. 2b) are poorly predicted by single residue perturbations averaged over many directions (maximum overlap values are 0.40 ± 0.08 by residue 90 using $\mathbf{G}_{net}$, 0.37 ± 0.06 by residue 14 using $\mathbf{G}_{10}$, 0.35 ± 0.06 by residue 112 using $\mathbf{G}_{20}$). This is in contrast to our previous work on ferric binding protein whereby similar responses were obtained regardless of the direction of perturbation of residues that control ligand entry-exit mechanisms [5]. Here, we also find that the length of the trajectory used in the calculations does not affect the results (compare $\mathbf{G}_{10}$ and $\mathbf{G}_{20}$ results in Fig. 2) implying that there are no major conformational changes occurring during the length of the simulation.

Although the conformational change is not well-predicted by average single-residue perturbations, this does not necessarily mean that one cannot invoke it via such perturbations. In fact, we find that when the directionality of the perturbation is considered separately, it is possible to recover the motion. Residues 27, 31, and 109 consistently emerge with an overlap in the range of 0.61–0.67 irrespective of the kernel used. These residues reside on an equivalent helix on either the N- or the C-lobe. Their topology is shown in fig. 1c, where the two domains are overlaid. Thus, the apo-form may be manipulated from this particular helix of the EF hand in either of the domains to generate the observed conformational change. The experimental displacement vectors of the conformational change (fig. 3a) and the response profile resulting from perturbation of residue 31 (fig. 3b) are compared in fig. 3c. It is observed that the response captures motion of both those in the C terminal lobe and the linker region, while in the N-lobe, the motion of the perturbed helix is mimicked (see fig.3c). The motion observed in fig. 3b, obtained by inserting the directed perturbation towards the outward normal of the helical curvature (shown in yellow), is mainly a twisting in the linker region and closing of the lobes. In

particular, the correlated nature of the motion in the linker implies its function as a transducer of binding information between the two lobes.
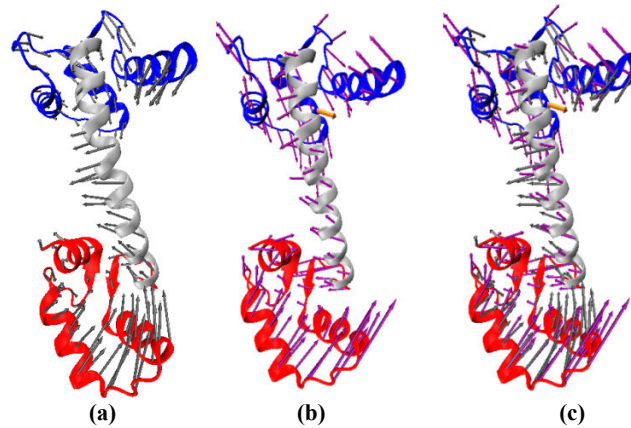

Figure 3. (a) Experimentally determined displacement vectors between the apo and holo forms, overlaid on apo CaM. **(b)** Response profile resulting from the perturbation of residues 31 in the direction shown in purple. Force vector applied to residue 31 is shown in orange. **(c)** Overlaid view of the vectors in (a) and (b); the OL between the two vectors is 0.67.

We use a simplified set of coordinates to capture the main features of this motion by reducing the structure of the protein to five points in space. These points are schematically shown in fig. 4a. Three of these points are the center of masses (COMs) of the N terminus (point 1), C terminus (point 5) and the linker (point 3). In addition, residues 64 and 93 are used to mark the end points of the linker (points 2 and 4, respectively). We then define two main degrees of freedom to capture the essence of the motions of the two lobes relative to each other. The bending angle θ defines the bending motion observed in the linker; φ defines the relative rotation of the two lobes around the linker as a torsional angle. The schematic representation of θ and φ are also displayed in fig. 4a. The θ and φ values computed throughout the 20 ns long MD trajectory for the apo-form are shown in fig. 4b. We find that the bending motion of the linker changes by $\pm 20°$ throughout this time window, and it essentially maintains the collinear arrangement of the two lobes ($<\theta> = 160°$). The torsional motion of the two lobes with respect to each other is more variable, changing by $\pm 40°$ in this time window; the average value of the torsion $<\varphi>$ is 70° so that the two lobes are nearly at right angles to each other in space. Thus, there are not any conformational changes occurring during the trajectory.
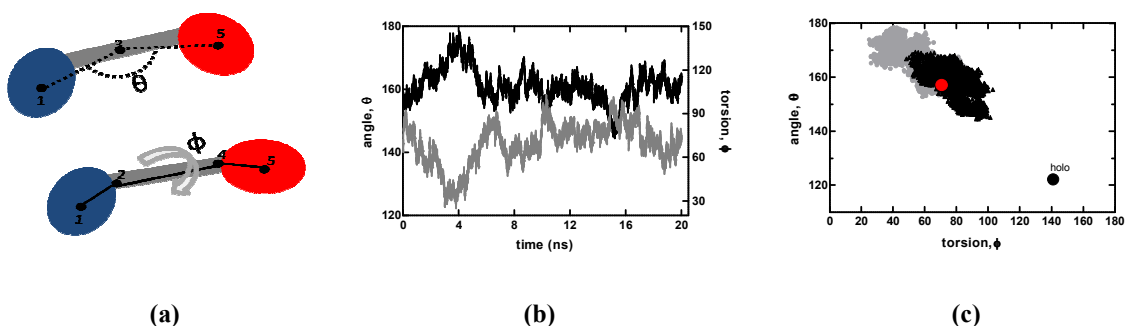

Figure 4. (a) Schematic display of the five reduced points, the bending angle θ and the torsion φ. C terminal is in red, N terminal is in blue. **(b)** MD trajectory of θ and φ. Left y-axis for θ (black) and right y-axis for φ (gray). **(c)** θ vs. φ plotted along the MD trajectory; first 10 ns in black, the additional 10 ns in gray. Red point shows the initial θ and φ values of apo form and black point shows those of holo form calculated from the PDB coordinates.

To locate the conformational distance of the holo form from the apo structure as well as the conformations sampled during the MD run, we plot θ vs φ values in fig. 4c. The initial apo structure, shown as the red point, resides in the midst of the values scanned during the MD trajectory, while the holo form (black point) is clearly located farther apart in conformational space. The 3D structure of the holo form is closed such that the bending of the two lobes is at 120°, which is now far from the collinear arrangement of the lobes. In addition, the two lobes have rotated by an additional 70° to a value of 140° which makes them face the opposite sides of the linker. According to this coarse-grained picture of the protein, the main motion upon ligand binding is the bending and the rotation of the two lobes around the linker axis.

CONCLUSIONS

PRS is a powerful method which can capture the conformational change between the apo and holo forms with an overlap of 0.65 via perturbing a single residue (31 or 109, residing in equivalent positions on the two lobes). These residues are located far from either the Ca^{+2} ion binding sites and the TFP binding region. Re-examining fig. 3b we find that these perturbations induce correlated motions in the linker so that the displacement vectors act in a helical direction, leading to the torsional motion quantified by φ in figure 4. In addition, the motion in the N-lobe is also correlated, resulting in a closing of this domain towards the C-lobe, quantified by θ as a bending motion. Our future work will focus on investigation of these sites as attractive allosteric control points for the overall conformational change.

REFERENCES

1. M. Ikura, J.B. Ames, *Proc. Natl. Acad. Sci.* **103**,1159-1164 (2006).
2. J.L. Fallon, F.A. Quicho, *Structure* **11**,1303-1307 (2003).
3. M. Vandonselaar, R.A. Hickie, W.J. Quail and T.J. Delbaere, *Struct. Biol.* **1**,795-801 (1994).
4. C.M. Shepherd and H.J. Vogel, *Biophys. J.* **87**,780–791 (2004).
5. C. Atilgan, A.R. Atilgan, *PLoS Comput. Biol.* **5**, e1000544 (2009).
6. M. Ikeguchi, J. Ueno, M. Sato, and A. Kidera, *Phys. Rev. Lett.* **94**, 078102 (2005).
7. L.S. Yilmaz, A.R. Atilgan, *J. Chem. Phys.* **113**, 4454–4464 (2000).
8. C. Atilgan, Z.N. Gerek, S.B. Ozkan, A.R. Atilgan, *Biophys. J.* **99**, 933-943 (2010).
9. A.R. Atilgan, S.R. Durell, R.L. Jernigan, M.C. Demirel, O .Keskin, et al., *Biophys. J.* **80**, 505–515 (2001).
10. C.Baysal, A.R. Atilgan, *Proteins.* **45**, 62–70 (2001).
11. C.Baysal, A.R. Atilgan, *Proteins.* **43**, 150–160 (2001).
12. Y.S.Babu, C.E. Bugg , W.J. Cook , *J.Mol.Biol.* **204**, 191-204 (1988).
13. W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **14**, 33–38(1996).
14. J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, et al., *J. Comput. Chem.* **26**,1781–1802 (2005).
15. B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, et al., *J. Comput. Chem.* **4**, 187–217 (1983).
16. T. Darden, L. Perera, L.P. Li, L. Pedersen, *Struc. Fold.Des.* **7**, R55–R60 (1999).