Driving Calmodulin Protein towards Conformational Shift by Changing Ionization States of Select Residues

Sunita Negi[1], Ali Rana Atilgan and Canan Atilgan
Sabanci University, Faculty of Engineering & Natural Science,
Tuzla, 34956 Istanbul, Turkey

E-mail: snegi@sabanciuniv.edu

Abstract. Proteins are complex systems made up of many conformational sub-states which are mainly determined by the folded structure. External factors such as solvent type, temperature, pH and ionic strength play a very important role in the conformations sampled by proteins. Here we study the conformational multiplicity of calmodulin (CaM) which is a protein that plays an important role in calcium signaling pathways in the eukaryotic cells. CaM can bind to a variety of other proteins or small organic compounds, and mediates different physiological processes by activating various enzymes. Binding of calcium ions and proteins or small organic molecules to CaM induces large conformational changes that are distinct to each interacting partner. In particular, we discuss the effect of pH variation on the conformations of CaM. By using the pKa values of the charged residues as a basis to assign protonation states, the conformational changes induced in CaM by reducing the pH are studied by molecular dynamics simulations. Our current view suggests that at high pH, barrier crossing to the compact form is prevented by repulsive electrostatic interactions between the two lobes. At reduced pH, not only is barrier crossing facilitated by protonation of residues, but also conformations which are on average more compact are attained. The latter are in accordance with the fluorescence resonance energy transfer experiment results of other workers. The key events leading to the conformational change from the open to the compact conformation are (i) formation of a salt bridge between the N-lobe and the linker, stabilizing their relative motions, (ii) bending of the C-lobe towards the N-lobe, leading to a lowering of the interaction energy between the two-lobes, (iii) formation of a hydrophobic patch between the two lobes, further stabilizing the bent conformation by reducing the entropic cost of the compact form, (iv) sharing of a Ca^{+2} ion between the two lobes.

1. Introduction

Calmodulin (CaM) is a protein which has been extensively studied due to its pivotal role as an intracellular calcium ion receptor as well as in the calcium signaling pathways in the eukaryotic cells [1]. CaM is observed to have different conformations influenced by the external environmental factors such as temperature, ionic strength and pH (see[2] and references cited therein). Furthermore, this protein can bind to a variety of other proteins or small organic compounds, and mediates different physiological processes by activating various enzymes [3, 4]. Binding of Ca^{+2} and proteins or small organic molecules to CaM induces large conformational changes that are distinct to each interacting partner [1, 4, 5].

---

[1] Corresponding author.

The plethora of data on CaM conformations come from diverse sources, including (i) "static" methods such as x-ray crystallography [6], (ii) dynamic methods that are based on averaging over an ensemble of structures, such as nuclear magnetic resonance (NMR) [7], and (iii) dynamic information obtained from single molecule spectroscopy approaches, such as fluorescence resonance energy transfer (FRET) [8]. There is plethora of data that identify CaM as a protein which samples multiple conformations. They also distinguish the shifts in the distribution of their populations by changing environmental conditions such as pH, Ca^{+2} concentrations, and ionic strength. None of these, however, are able to directly identify the exact conformations that are sampled by the protein. They also fall short of directly describing the events that trigger the changes between different conformational states, which occur on millisecond time scales.

Experimentally, the conformation distributions of CaM are studied by FRET under various environmental conditions[8]. Of particular interest to us is that compaction of the protein was observed by lowering the pH from 7.4 to 5.0 [9]. The main drawback with these experimental studies is that they neither directly identify the exact conformations sampled by CaM, nor hint at the events that trigger the conformational changes[9]. Here, we aim to study the conformational multiplicity of CaM at reduced pH.

Computationally, Donnini et al. [10] present a method, implemented within molecular dynamics (MD) package GROMACS [11], for constant pH MD simulations in explicit solvent that is based on the λ-dynamics approach. Therein, the dynamics of the titration coordinate λ, which interpolates between the protonated and deprotonated states, is driven by generalized forces. Lee et al[12] introduced another method for performing constant-pH molecular dynamics (PHMD) simulations utilizing the generalized Born implicit solvent model. This approach employs an extended Hamiltonian in which continuous titration coordinates propagate simultaneously with the atomic motions of the system.

In this manuscript, we discuss the effect of pH variation on the conformations of calcium-loaded CaM by a different approach. The pKa values of the charged residues are determined; those charged residues with values shifted into the physiological range are determined. As a result, the protonation states of nine residues are changed to mimic the overall electrostatics of the protein at pH 5.0. MD simulations of length 190 ns are carried out by assigning these new protonation states. We find that CaM undergoes a conformational shift that leads to a more compact form at a lower pH value. We describe the key events that lead to the final compact conformation.

2.   Computational Model: Molecular Dynamics

CaM consists of 148 amino acid residues, consisting of three domains, namely the N-lobe, the C-lobe and the central linker/helix as shown in Fig.1(a) [residues 5-68:N lobe, residues 69-91:linker, residues 92-147:C lobe]. There exist two forms of the CaM protein whereby apo-CaM is free of calcium ions, while holo-CaM is fully loaded with four calcium ions. In this study we focus on holo-CaM which is commonly observed to interact with other proteins[13]. Fully loaded CaM free of other ligands has two distinct experimentally determined conformations: In many x-ray structures, CaM has an open form with a dumbbell shape containing two domains, joined by an extended linker. It is exemplified by the protein data bank (PDB [14]) structure 3CLN. Another form is compact, whereby the central linker is bent (PDB code 1PRW). Each of the lobes holds two calcium ions (shown as grey spheres in figure 1) in their respective EF-hand motifs. Upon binding a ligand, the domains move relative to each other and the flexible linker region changes conformation accordingly.

We use the PDB structure coded 3CLN as the initial conformation. The data for the residues 1-4 and 148 are not included in this structure as the X-ray data for these residues are not available. We use the NAMD package to model the dynamics of the protein-water system[15]. The protein is soaked in a water box forming a layer of at least 10 Å measured from any residue. The protonation of the residues are

determined such that we mimic the most probable state of each ionizable residue at pH 5. Details on the parameters of a similar simulation at pH 7.4 may be found in[2] where results from 120 ns long MD simulations are reported. For the reduced pH environment, we protonate nine residues as described next.

(a)   (b)
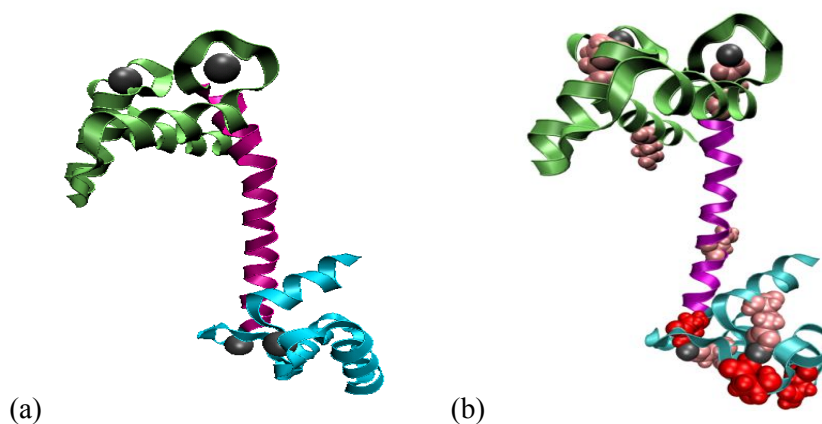
Figure 1 (a) Fully loaded Calmodulin (CaM) with four calciums; N-lobe shown in green, the linker in purple and the C-lobe in cyan. (b) Calmodulin shown with the nine residues which are protonated so as to mimic the most probable ionization states at pH 5 (GLU : pink, ASP: red).

Four different servers H++ [16, 17], pKd[18, 19], propKa 3.1[20] and PHEMTO[21] are used to calculate the pKa of all the titratable residues. There are a total of 52 such sites in CaM, two of which are Tyr with high pKa values, and one His which we take to be uncharged. The standard error on the mean of the remaining Glu/Asp/Arg/Lys residues is calculated to be less than 0.5 units in 34 cases, and less than 1.3 units for 44 cases. In the remaining five, the residues are either consistently protonated above pH 5 (e.g. for residue Glu 67, we have pKa = 11.2, 5.4, 13.4 and 8.5, respectively for H++, pKd, propKa 3.1 and PHEMTO) or they are deprotonated for at least three of the four pKa predictors. This lets us assign the protonation states consistently. We therefore find that the acidic residues 11, 31, 67, 84, 93, 104, 122, 133 and 140 have upshifted pKas to 5.5 or more, compared to the standard values of ca. 4. Out of these nine residues 11, 31, 67 are on the N-lobe, 84 is on the central linker and the remaining five residues are on the C-lobe; GLU and ASP residues are shown in pink and red respectively in Fig.1(b). All other negatively charged residues are observed to be near their standard pKa values of approximately 4 in the protein environment, so that they are kept in the charged state throughout the simulation. None of the positively charged residues have downshifted values, and their protonation states are also unaltered.

The system is solvated using the VMD 1.8.7 program with solvate plug-in version 1.2[22]. The CharmM27 force field parameters are used for protein and water molecules[23]. Water molecules are described by the TIP3P model. The initial box has dimensions 72×92×72 Å containing 31660 atoms neutralized by standard addition of Na$^+$ and Cl$^-$ ions. Long range electrostatic interactions are calculated by the particle mesh Ewald method[24], with a cutoff distance of 12 Å and a switching function at 10 Å. RATTLE algorithm[25] is applied to use a step size of 2 fs in the Verlet algorithm. Temperature control is carried out by Langevin dynamics with a dampening coefficient of 5/ps. Pressure control is attained by a Langevin piston. Volumetric fluctuations are preset to be isotropic. The system is run in the NPT ensemble at 1 atm and 310 K until volumetric fluctuations are stable to maintain the desired average pressure. The run in the NPT ensemble is extended to a total of 190 ns. The coordinate sets are saved at 2 ps intervals for subsequent analysis, leading to $T = 95000$ snapshots.

We note that we carried out two separate simulations. The first one is of length 190 ns, where all results reported here are calculated from. The second one is 100 ns long, and is used to confirm that the main features observed in the first trajectory are repeatable. We found that the same conclusions are reached from both sets of simulations.

3.   Results and Discussion

### 3.1 Conformation change obtained at pH 5.0

As discussed above, we protonate nine residues so as to mimic their most probable states at pH 5 and perform extensive MD simulations. For a comparison with the FRET experiment results[9], we monitor the distance between residues 34 and 110 in CaM throughout the simulation. The cumulative distribution of separation distance between residues 34 and 110 is shown in figure 2, as the simulation proceeds. On average, the distance shrinks with time, implying that there is a change in the conformation of the protein during the simulation.
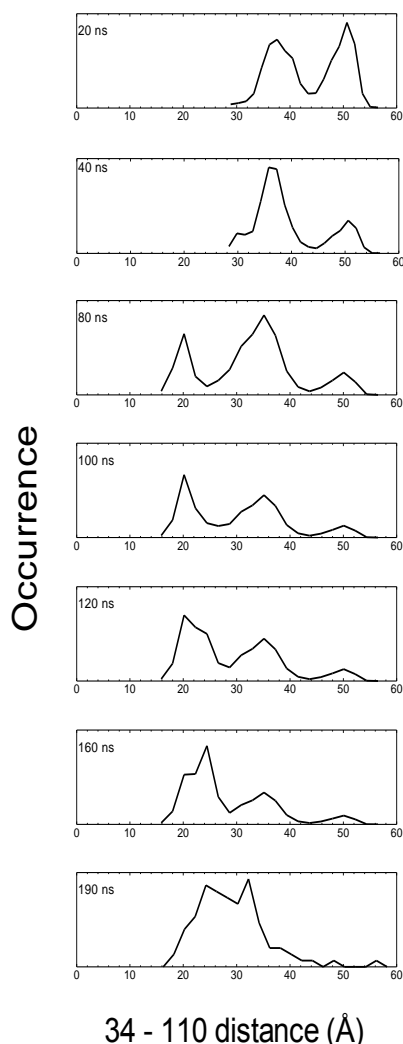
Figure 2 The distribution of the distance (in Å) between residues 34 and 110, evaluated cumulatively until the indicated instants of the simulation.

The root mean square distances (RMSDs) of the different parts of the protein are calculated with respect to the original structure and is presented in figure 3. While the overall conformation of the

chain shows abrupt changes in two steps, first at ca. 15 ns, then at ca. 65 ns, the rearrangements within the different units are relatively small. Thus, the main conformational change is due to the relative arrangement of the units with respect to each other. Inspecting this figure more closely, we further find that the C-lobe motions are relatively larger, while the linker maintains its original helical structure. The latter point is important because various authors have argued that the loss of the helicity in the central linker might be the source of the conformational multiplicity of CaM [26-28].
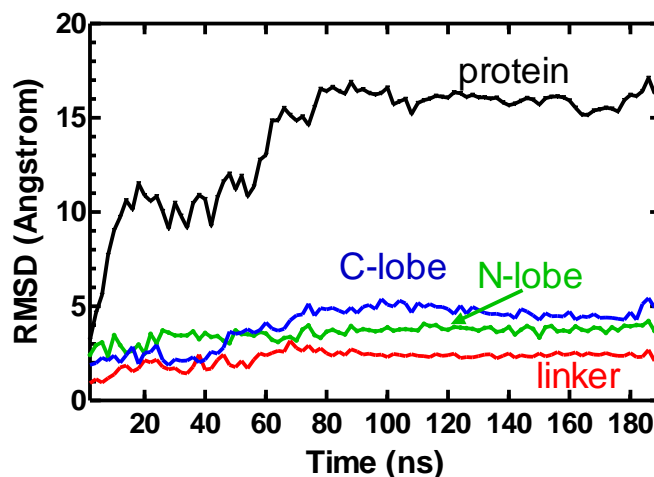
Figure 3 RMSD as a function of time at PH 5.0. RMSD of C-lobe is observed to be higher than that of N-lobe and linker

To capture the main features of the relative motions of the two lobes and the linker, we reduce the protein structure to five points in space. These points are schematically shown at the top of figure 4. 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 69 and 91 are used to mark the beginning and end points of the linker (points 2 and 4). We then define two main degrees of freedom to capture the essence of the motions of the two lobes relative to each other. θ angle defines the bending motion observed between the N and C lobes, while φ angle defines the relative rotation of the two lobes around the linker as a virtual dihedral angle. Similar virtual dihedral angle definitions on CaM were previously made[29].

As seen in figure 4 (bottom), during the first 40 ns, the bending angle covers a relatively restricted region of values in the range 120-180°, while the torsion angle is more flexible, with values in the range 60-140°. In our previous work, where all nine acidic residues are in a deprotonated state, this range of values are observed for the whole 120 ns of the trajectory[2], while extension of those simulations further to 200 ns kept the same sampling range (data not shown). Thus, there is normally a barrier to bending of the two lobes with respect to each other. However, in this work, starting at ca. 55 ns, the bending motion is realized. At ca. 80 ns, the bending angle obtains a value near 110° implying a closer association of the N- and C-lobes. By 100 ns, the bending angle is observed to attain values as small as 90°. For the last 70 ns, the protein is observed to have come to a stable point where the (θ, φ) is $(100\pm10°,100\pm20°)$.
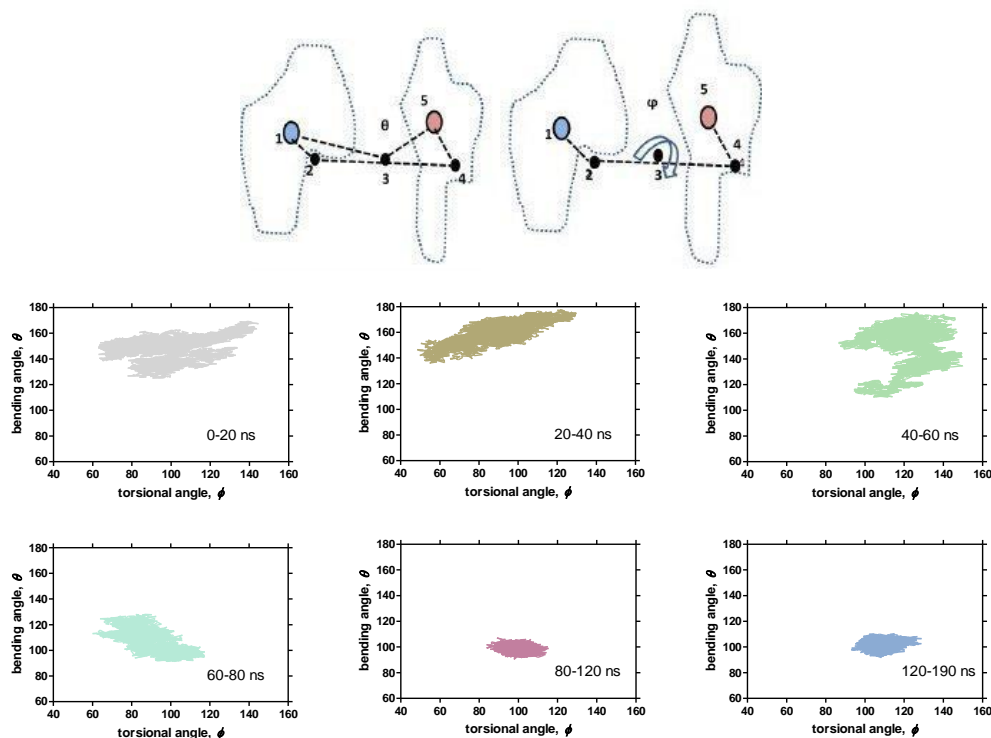
Figure 4 (top) Schematic representations of the defined bending angle θ and torsional angle φ. (bottom) (θ, φ) values covered in different parts of the simulation. A stable configuration is attained by the protein for the last 120 ns.

3.2 Shifts in the pKa values of the residues

The shift in the pKa value of a titratable residue depends on the change in its local environment, reflected as free energy differences. Therefore, large conformational shifts may substantially change the pKa values of these residues. While monitoring the key events taking place during the simulation, we also calculate the pKa values of all the charged residues. The trends for selected residues are shown in figure 5. Residues 11, 84 and 139 are observed to show a monotonic increase in their pKa values during the simulation. Note that of these, 11 and 84 are amongst the protonated residues in this MD simulation, while 139 is not. Other than these three residues, there are none which show a significant change in their pKa value during the simulation.
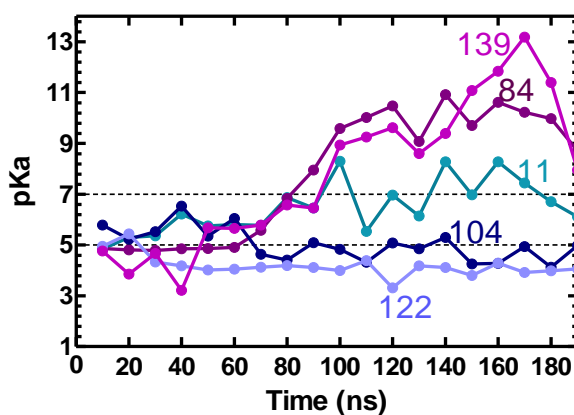
Figure 5 Changes in the pKa values of selected residues during the simulation, calculated using the propKa 3.1 server.

3.3 Interaction energy between the n and c

Using the "NAMD energy" plugin in VMD, we calculate the interaction energy between the N- and the C-lobes as shown in the figure 6. The electrostatic interactions are observed to make the major contribution to the non-bonded interaction energy. A dip is observed in the energy around 40 ns followed by two larger dips at 76 ns and 120 ns. The van der Waals part of the energy is observed to be an order of magnitude smaller. After 120 ns, the energy is observed to increase for ca. 20 ns, and stabilizes thereafter. We seek to associate the main events that take place during the conformational change with the inter-lobe interactions next. Since the overall stability of the system is determined by the free energy, we conjecture that entropic contributions must also play a role to compensate for the raise in the interaction energy along the trajectory.
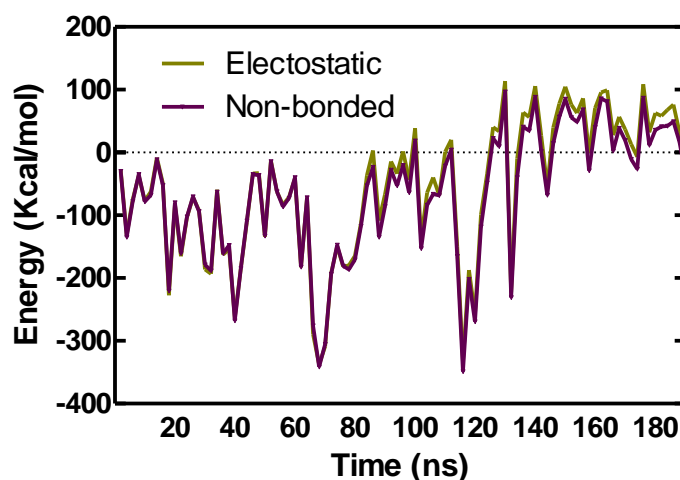

Figure 6 The non-bonded interaction energy between N and the C-lobes throughout the MD simulation; the electrostatic contribution is also shown, displaying that the van der Waals term has negligible effect on the conformational change.

3.4 Key events leading to the conformational change

In order to provide a more detailed understanding of the observed conformational changes, we study all the possible interactions between residues in the protein throughout the simulation. A salt bridge formation is observed at 40 ns between the residues 74 (ARG) of the linker and 7 (GLU) of the N-lobe as shown in figure 7a. This salt bridge restricts the motions of the N-lobe, and its formation causes a dip in the energy of the system. Once formed, it is present in the rest of the simulation. At 67 ns, the calcium ion in the third EF-hand motif, which is located in the C-lobe immediately following the linker, moves slightly away as displayed in figure 7b. This event is immediately followed by the partial unfolding of the EF-hand motif, causing a bending motion in the rest of the C-lobe. The major cause of this event is that the protonation of residue 104, which is part of this particular EF-hand motif, weakens the grip of the motif on the Ca^{+2}. This is also the reason why the C-lobe attains a higher RMSD value than both the linker and the N-lobe from this point onward (recall figure 3). The overall result of the loosened motif is the association of the C-lobe with the N-lobe which causes a dip in the energy around this point.

As shown in figure 7c, a hydrophobic patch is fully formed at 80 ns which helps stabilize the system; since hydrophobicity has entropic contribution to the free energy, this patch also compensates for the increase in the interaction energy of the two lobes, observed from 67 ns to 80 ns. Close to 120 ns, the calcium ion of the EF-hand motif 4 comes and sits in the center of the two lobes as shown in figure 7d. By 156 ns, this calcium is observed to keep this position at the center of the N and the C lobes, stabilized by two negatively charged residues 84 and 139. The linker is retains its helical structure throughout the simulation. This form of the linker is in contrast to those observed in other ligand-

bound, compact forms of CaM where the linker is bent into two at position 74[29]. In order to monitor the stochasticity of our MD simulation, we repeat this simulation for 100 ns with a different random number seed and observe a similar sequence of events: A conformational change from the open to the closed form occurs at 90 ns, preceded by a salt bridge formation between the linker and the N-lobe at 64 ns and followed by the hydrophobic patch formation at 92 ns.

(a) 40ns

(b) 67 ns

(c) 80 ns

(d) 120 ns
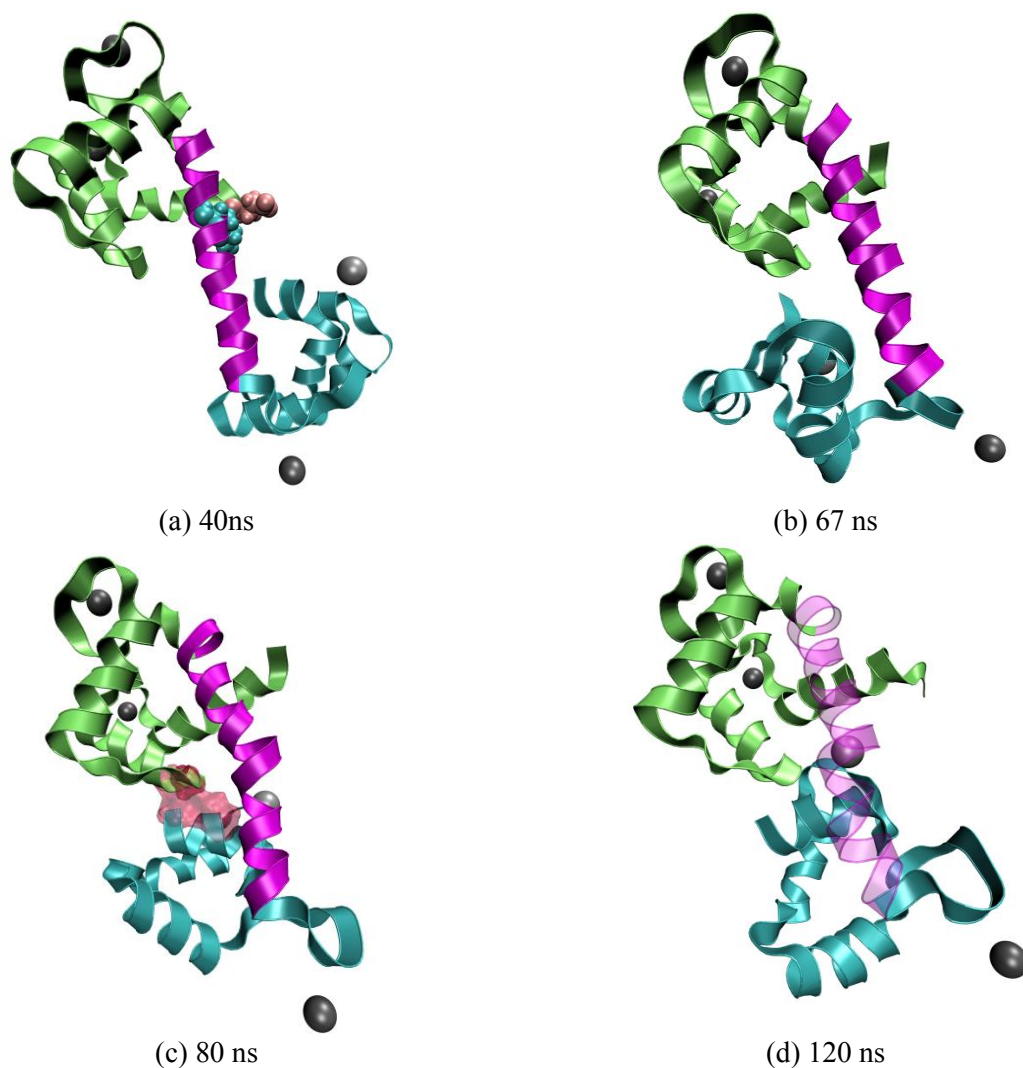
Figure 7**.** Snapshots from the MD simulation; N-lobe is in green, linker is in purple, C-lobe is in cyan. At (a) 40 ns, a salt bridge is formed between residues 74 (ARG) in the linker and 7 (GLU) in the N-lobe, these residues are shown in space-filling representation; (b) 67 ns, a calcium from the N-lobe is observed to move away from its EF-hand motif, this event is immediately followed by an increased mobility in the associated EF-hand motif; (c) 80 ns, a hydrophobic patch is formed on the other side of the protein shown by the new space filling atoms adjoining the two lobes; (d) 120 ns, a calcium comes and sits in the center of the two lobes, stabilized by negatively charged residues from each lobe. This form is kept in the rest of the simulation.

4.   Summary
Fully loaded CaM is studied by MD simulations under conditions mimicking the most probable protonation states of residues at pH 5. By lowering the pH, local conformational changes in the protein are induced, leading to shifts in the energy landscape. These pave the way to the compact forms reminiscent of those detected by indirect experimental methods (compare figure 2 to figure 3 in ref[9]). A salt bridge and a hydrophobic patch are observed to be formed as key events leading to the

conformational change by stabilizing the system. This conformation change in fully-loaded CaM is observed in two independent simulations started with different random number seeds.

Our current view suggests that at high pH, barrier crossing to the compact form is prevented by the repulsive electrostatic interactions between the two lobes. Thus, the time scale of jumps between open-closed forms of CaM is measured on the order of 100 μs. Lowering of pH significantly reduces the barrier height, while possibly also changing the overall conformational landscape.

Adaptation to different pH values in various subcellular compartments is thought to be directly related to protein stability[30] and pH of optimal binding affinity of interacting proteins[31]. Methods to determine how proteins adapt to cellular and sub-cellular pH are currently sought-after[32]. These findings lend clues as to how the conformations of proteins and therefore their associated functions may be modulated in the differentially controlled pH of the different compartments of the cell. The findings in the current study must be supported by additional MD simulations whereby the protonation states of not all but only subsets of the nine residues acted upon in this work are altered. Our work in this direction is under way.

5.   References
1.   M. Ikura and J. B. Ames, *Proceedings of the National Academy of Sciences of the United States of America*, 2006, **103**, 1159-1164.
2.   A. R. Atilgan, A. O. Aykut and C. Atilgan, *Journal of Chemical Physics*, 2011, **135**, 15.
3.   J. L. Fallon and F. A. Quiocho, *Structure*, 2003, **11**, 1303-U1307.
4.   M. Vandonselaar, R. A. Hickie, J. W. Quail and L. T. J. Delbaere, *Nature Structural Biology*, 1994, **1**, 795-801.
5.   J. P. Junker and M. Rief, *Proceedings of the National Academy of Sciences of the United States of America*, 2009, **106**, 14361-14366.
6.   Y. S. Babu, C. E. Bugg and W. J. Cook, *Journal of Molecular Biology*, 1988, **204**, 191-204.
7.   I. Bertini, C. Del Bianco, I. Gelis, N. Katsaros, C. Luchinat, G. Parigi, M. Peana, A. Provenzani and M. A. Zoroddu, *Proceedings of the National Academy of Sciences of the United States of America*, 2004, **101**, 6841-6846.
8.   B. Slaughter, M. Allen, J. Unruh, R. Urbauer and C. Johnson, *Journal of Physical Chemistry B*, 2004, **108**, 10388-10397.
9.   B. D. Slaughter, J. R. Unruh, M. W. Allen, R. J. B. Urbauer and C. K. Johnson, *Biochemistry*, 2005, **44**, 3694-3707.
10.   S. Donnini, F. Tegeler, G. Groenhof and H. Grubmuller, *Journal of Chemical Theory and Computation*, 2011, **7**, 1962-1978.
11.   E. Lindahl, B. Hess and D. van der Spoel, *Journal of Molecular Modeling*, 2001, **7**, 306-317.
12.   M. S. Lee, F. R. Salsbury and C. L. Brooks, *Proteins-Structure Function and Bioinformatics*, 2004, **56**, 738-752.
13.   A. Crivici and M. Ikura, *Annuual Review of Biophysics and Biomolecular Structure*, 1995, **24**, 85-116.
14.   H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Research*, 2000, **28**, 235-242.

15. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *Journal of Computational Chemistry*, 2005, **26**, 1781-1802.
16. J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath and A. Onufriev, *Nucleic Acids Research*, 2005, **33**, W368-W371.
17. J. Myers, G. Grothaus, S. Narayanan and A. Onufriev, *Proteins* 2006, **63**, 928-938.
18. B. M. Tynan-Connolly and J. E. Nielsen, *Nucleic Acids Research*, 2006, **34**, W48-W51.
19. T. Connolly, B.M. and J. Nielsen, *Protein Science*, 2007, **16**, 239-249.
20. M. Rostkowski, M. H. M. Olsson, C. R. Sondergaard and J. H. Jensen, *Bmc Structural Biology*, 2011, **11**, 6.
21. A. A. Kantardjiev and B. P. Atanasov, *Nucleic Acids Research*, 2009, **37**, W422-W427.
22. W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics*, 1996, **14**, 33.
23. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *Journal of Computational Chemistry*, 1983, **4**, 187-217.
24. T. Darden, L. Perera, L. P. Li and L. Pedersen, *Structure with Folding & Design*, 1999, **7**, R55-R60.
25. H. C. Andersen, *Journal of Computational Physics*, 1983, **52**, 24-34.
26. G. Fiorin, R. R. Biekofsky, A. Pastore and P. Carloni, *Proteins-Structure Function and Bioinformatics*, 2005, **61**, 829-839.
27. A. Likić, R. Gooley, P. Speed and E. Strehler, *Protein Science*, 2005, **14**, 2955-2963.
28. C. M. Shepherd and H. J. Vogel, *Biophysical Journal*, 2004, **87**, 780-791.
29. W. Wriggers, E. Mehler, F. Pitici, H. Weinstein and K. Schulten, *Biophysical Journal*, 1998, **74**, 1622-1639.
30. P. Chan and J. Warwicker, *BMC Biology*, 2009, **7**, 69.
31. Z. Zhang, S. Witham and E. Alexov, *Physical Biology*, 2011, **8**, 3.
32. B. Garcia-Moreno, *Journal of Biology*, 2009, **8**, 98.91-98.94.