

Crystal structure of *Arabidopsis thaliana* calmodulin7 and insight into its mode of DNA binding

Sanjeev Kumar¹, Mohit Mazumder¹, Nisha Gupta², Sudip Chattopadhyay² and Samudrala Gourinath¹

¹ School of Life Sciences, Jawaharlal Nehru University, New Delhi, India

² Department of Biotechnology, National Institute of Technology, Durgapur, India

Correspondence

S. Gourinath, School of Life Sciences, Jawaharlal Nehru University, New Delhi 110067, India

Fax: +1 91-11-26742916/2558

Tel: +1 91-11-26704513

E-mail: sgourinath@mail.jnu.ac.in

(Received 14 May 2016, revised 6 July 2016, accepted 1 August 2016, available online 24 August 2016)

doi:10.1002/1873-3468.12349

Edited by Richard Cogdell

Calmodulin (CaM) is a Ca^{2+} sensor that participates in several cellular signaling cascades by interacting with various targets, including DNA. It has been shown that *Arabidopsis thaliana* CaM7 (AtCaM7) interacts with Z-box DNA and functions as a transcription factor [Kushwaha R *et al.* (2008) *Plant Cell* 20, 1747–1759; Abbas N *et al.* (2014) *Plant Cell* 26, 1036–1052]. The crystal structure of AtCaM7, and a model of the AtCaM7-Z-box complex suggest that Arg-127 determines the DNA-binding ability by forming crucial interactions with the guanine base. We validated the model using biolayer interferometry, which confirmed that AtCaM7 interacts with Z-box DNA with high affinity. In contrast, the AtCaM2/3/5 isoform does not show any binding, although it differs from AtCaM7 by only a single residue.

Keywords: CaM; molecular modeling; protein crystallization; protein–DNA interaction

Calmodulin (CaM) are ubiquitous eukaryotic proteins that can bind to a variety of protein targets in response to Ca^{2+} signals. CaM plays essential role in Ca^{2+} signaling, regulating numerous intracellular processes such as cell motility, growth, proliferation, and apoptosis [1]. CaM binds Ca^{2+} ions using its helix-loop-helix (EF-hand) structural motif, and this motif generally undergoes large conformational changes upon Ca^{2+} binding [2]. In the Ca^{2+} -loaded form, the CaM adopts stable state, and each EF-hand opens so that its two alpha helices become perpendicular to each other. In the Ca^{2+} -free (apo) form, CaM adopts a closed and flexible state where EF-hand motifs are in closed conformation [3–5]. In contrast to apo form, Ca^{2+} -loaded CaM binds to many (> 300) target proteins that regulates the various biological processes [6–8].

CaM responds to a wide range of Ca^{2+} concentrations (10^{-12} to 10^{-6} M) in Ca^{2+} -dependent signal transduction, after binding of Ca^{2+} to EF hand

motifs, CaM can bind to different target proteins to accomplish these physiological roles [9,10]. CaMs are found to be involved in various signaling event and these signaling are governed in Ca^{2+} -dependent as well as Ca^{2+} -independent manners. In many cases these signaling mechanisms have been elucidated on structural basis in from of protein–protein (CaM complex with its target proteins) complex structures [11,12]. CaM can regulate basic helix-loop-helix transcription factors where CaM inhibits DNA–protein interactions by competing with the DNA-binding domains of the basic helix-loop-helix proteins [13]. Helix-loop-helix motifs, such as that in the EF-hand-containing protein DREM, have been reported to interact with DNA [14–16]. Aside from interactions with various proteins, CaM can also interact with DNA and serve as transcription factors.

The *A. thaliana* genome contains seven CaM genes that encode four protein isoforms: CaM1/CaM4,

Abbreviations

BLI, biolayer interferometry; CaM, calmodulin.

CaM2/3/5 (AtCaM2/3/5), CaM6, and CaM7 (AtCaM7). *A. thaliana* CaM7 (AtCaM7) has been reported to be a transcriptional regulator that directly interacts with promoters of light-inducible genes and promotes photomorphogenesis. AtCaM7 has been specifically shown to interact with Z/G-box of light-regulated CAB1 and RBCS-1A minimal promoters [17] and its other isoform, CaM2/3/5, does not interact with the Z/G-box, although it is different from AtCaM7 by a single residue (AtCaM7 has Arginine at 127th position, whereas AtCaM2/3/5 has Lysine). Recently it has also been shown that AtCaM7 and the transcription factor HY5 directly interact with the HY5 promoter to mediate the transcriptional activity of HY5 during *A. thaliana* seedling development, and hence promote photomorphogenesis [18]. While there have been several reports related to CaM-target protein complexes and on their different modes of binding, there have been no published reports delineating the mode of CaM–DNA interaction. Moreover, no atomic structure of CaM in complex with DNA is available. In this study, we are first time reporting the crystal structure of *A. thaliana* CaM7 and its interaction study with Z-box DNA by using bilayer interferometry (BLI) experiment. Based on the interaction study, molecular dynamics simulation and molecular modeling, we are proposing a theoretical complex model of AtCaM7–Z-box DNA, hence providing for the first time an explanation of how a CaM may bind with DNA.

Materials and methods

Overexpression and purification of AtCaM7

The plasmid containing *Arabidopsis thaliana* CaM7 was transformed into *E. coli* BL21 (DE3) cells and protein was overexpressed and purified as described previously [17].

Crystallization of AtCaM7

The purified protein was concentrated to 20 mg·mL⁻¹ in 50 mM Tris pH 8.0 buffer containing 10 mM CaCl₂. Initial crystallization trials were carried out by using the hanging drop vapor diffusion method in 24-well plates with 2 μ L of protein solution mixed with an equal volume of precipitant solution and equilibrated against 500 μ L of precipitant containing 60–65% MPD, 5 mM CaCl₂, and 100 mM sodium acetate buffer pH 3.9–4.3. Crystallization plates were kept at 4 °C for equilibration. In initial trials, needle-shaped crystals appeared after approximately 12 h of equilibration. Diffraction quality crystals were obtained in subsequent trials by varying the protein concentration.

Data collection, processing, and structure determination

Crystals were soaked in cryoprotectant solution consisting of 65% MPD, 5 mM CaCl₂, and sodium acetate buffer pH 3.9. Single crystals were picked up in cryoloops and flash cooled in liquid nitrogen. X-ray diffraction data were collected at DBT BM-14 at ESRF. Crystal diffracted to a resolution of 2.2 Å. Diffraction data were processed and scaled using HKL-2000 [19]. The crystal belonged to space group P2₁2₁2₁ with unit cell parameters $a = 24.45$, $b = 67.90$, $c = 111.29$ Å. The initial structure was determined using molecular replacement with the MOLREP program [20] using potato CaM [21] (PDB entry 1RFJ) as the search model, as potato CaM has the highest sequence identity with *A. thaliana* CaM7 among the known structures in PDB. Four calcium atoms not included in the search model were identified in the subsequent electron density map (both in Fo-Fc and 2Fo-Fc), and these atoms correspond to the four EF-hands motifs of CaM and were included in the refinement. The final structure was obtained after iterative model building using the COOT graphics package [22] and refinement was carried out using REFMAC5 software [23] in the CCP4 suite. For this final model, the R_{work} and R_{free} are 19.0%, 23.8%, respectively. The structure has good stereochemistry as indicated by the program PROCHECK [24] with 97.6% of the residues lying in the most favored regions of the Ramachandran plot. The refined model of AtCaM7 and structure factors was deposited in the Protein Data Bank under the accession code 5A2H. The data collection and final refinement statistics are shown in Table 1.

Biolayer interferometry

To detect AtCaM7–Z-box interaction, bilayer interferometry (BLI) experiment was performed on fortbio octet K2 two-channel system. Both proteins AtCaM7 and AtCaM2/3/5 were immobilized on Ni-coated biosensor tips, sensors were prewet for 10 min before immobilization. Then, tips were loaded with 20 μ g of CaM7 and CaM2/3/5 for immobilization, till each sensor capture level reaches up to 0.5 nm. Analyte (double-stranded Z-box DNA 5'-ATCTATTCGTATACGTGTCACATCTATTCGTATACGTGTCAC-3') was diluted in the above-mentioned buffer, in various concentration of 62.5, 125, 250, 500 nM. Then, association (kon) and dissociation (koff) were monitored by dipping the biosensors in various concentrations of Z-box DNA samples, distributed in 96-microwell plates at a volume of 200 μ L per well. Binding study was performed in buffer containing HEPES pH 7.4, 0.02 tween20, and 1 mM CaCl₂. Signal changes from sensor tip loaded with ligand AtCaM7 and AtCaM2/3/5 (Incubated with Z-box DNA analyte) were subtracted with the reference data, binding of AtCaM7 with Z-box DNA was shown as 'nm' shift.

Table 1. Crystallographic data and refinement statistics.

X-ray source	ESRF DBT BM-14
Wavelength (Å)	0.977
Space group	P2 ₁ 2 ₁ 2 ₁
Unit cell parameter (Å)	
<i>a</i>	24.5
<i>b</i>	67.9
<i>c</i>	111.3
Rsym (%)	7.2 (19.9)
Completeness (%)	98.6 (94.2)
Total reflections	52934
Unique reflections	8910
Redundancy	5.9 (4.6)
Average I/S	27.9 (6.8)
Crystal mosaicity (°)	1.5
CC1/2	0.990 (0.962)
Refinement	
Resolution (Å)	50–2.2
<i>R</i> factor (%)	19.0 (27.0)
Free <i>R</i> factor (%)	23.8 (32.1)
Mean <i>B</i> factor (Å ²)	48.5
No. of atoms	
Protein/Ca/water/MPD/ACT	1157/4/48/3/1
Bond length (Å)	0.015
Bond angles (°)	1.66
Dihedral angles (°)	16.41
Cross validation error	0.214
Ramachandran plot statistics	
Favored region (%)	91.7
Additional allowed region (%)	8.3
Generously allowed region (%)	0.0
Outliers (%)	0.0

Statistics for the highest resolution shell are given in parentheses.

Binding were summarized as KD which was calculated from 'KD = koff/kon', where kon is the 'dissociation or off rate' and koff is the 'association or on rate'. Data were processed and analyzed using the OCTET data analysis software version 7.0 (ForteBio, Menlo Park, CA, USA) by a global fitting model assuming reversible binding and the experiment was performed at 30 °C temperature.

Molecular dynamics simulations

The crystal structure of AtCaM7 was used as the starting structure to obtain the AtCaM2/3/5 isoform using the ROSETTA BACKRUB server [25]. The final structure was selected on the basis of the lowest score (the first structure in the assembly of the lowest energy conformers) sampled by Monte Carlo simulated annealing using the Rosetta all-atom force field.

To study the difference in the DNA-binding properties of AtCaM7 and its isoforms (AtCaM2/3/5), both the systems were subjected to molecular dynamics simulations. Lennard–Jones parameters of the Ca²⁺ ions with a Vander Waals radius *R* of 1.7131 Å were obtained by xleap module

of the AMBER suite [26] for simulation. AMBERff99SB force field was used for the protein molecule. The input files for energy minimization, dynamics, and analysis were prepared with xleap. Both systems proteins were solvated using atomistic TIP3P water in a box with edges at least 12 Å from the complex. All simulations were performed using the AMBER molecular dynamics suite version 12. Energy minimization was first conducted with the steepest descent method and then switched to conjugate gradient every 500 steps for a total of 5000 steps with 0.1 kcal·(mol⁻¹·Å⁻²) restraints on all atoms of the complexes. Following this step, another two rounds of energy minimization were performed by only restraining the protein and further releasing all the restraints for 2000 steps of each round. Long-range Coulombic interactions were handled using the particle mesh Ewald (PME) summation. For the equilibration and subsequent production runs, the SHAKE algorithm was employed on all atoms covalently bonded to a hydrogen atom, allowing for an integration time step of 2 fs. The system was gently annealed from 0 to 300 K over a period of 500 ps using a Langevin thermostat with a coupling coefficient of 1.0 and 500 ps of density equilibration with weak restraints. The system was again equilibrated for 5 ns without any restraints. The production phase of the simulations was run without any restraints for a total of 150 ns on each system. Coordinates and energy values were collected every 1 ps throughout the simulations. Structural superimpositions of average structures of AtCaM7 and AtCaM2/3/5 were performed using the RAPIDO server. The images were prepared using Pymol [27].

Molecular docking simulations

The predicted structure of the AtCaM7-DNA complex was obtained using a series of modeling protocols. In the first step, the B-DNA (ATCTATTCGTATACGTGTCAC) was built using 3D-Dart service and using the HADDOCK server [28]. Furthermore, with the knowledge of the binding regions Z-box of DNA and Arg127 of AtCaM7 as the interacting residue, molecular docking simulations were performed using the HADDOCK WENMR GRID enabled docking server [29]. All HADDOCK runs were performed with 1000 structures for rigid body docking using backbone rmsd cut-off for clustering equal to 10.0 Å, and a minimum cluster size equal to 10. The final model was then further refined using FIREDOCK [30] and then subjected to AMBER-based energy minimization using steepest descent and conjugate gradient method. Prior to the mutational analyses, both the AtCaM7 and DNA structures were repaired with the FOLDX program [31] to remove potential steric clashes. FOLDX utilizes an empirical force field model to estimate the stability of a protein (ΔG wild-type in kcal·mol⁻¹) and to estimate the free energy difference (stability change, expressed as $\Delta\Delta G$) upon mutagenesis from wild-type for the mutations.

Results

Overall structure of *A. thaliana* CaM7

Sequence alignment with various CaM suggests that AtCaM7 shows maximum sequence homology with Potato CaM (Fig. 1A). Structure of AtCaM7 was solved by using structure of Potato CaM (PDB code 1RFJ) as a search model (discussed in detail in Materials and methods section). *A. thaliana* CaM7 crystallized with one molecule per asymmetric unit, along with four Ca²⁺ ions, three MPD molecules and 48 water molecules. The final refined structure includes 147 residues and is similar to other CaM-like structures. It has a dumbbell-like structure consisting of an N-terminal globular domain and a C-terminal globular domain, connected by a long central alpha helix. Similar to other CaM molecules, AtCaM7 contains seven α -helices, which are connected by loops that together form four EF-hand motifs (Fig. 1A) and each EF-hand motif is bound to one Ca²⁺ ion. Each Ca²⁺ is coordinated five residues, with glutamate at 12th position which provides two oxygen atoms for Ca²⁺ coordination (Fig. 1B). One water molecule is also involved in each case to complete a pentagonal bipyramidal coordination sphere.

Comparison with various CaM structure

Structural alignment of AtCaM7 suggest that it is very similar to potato CaM [21] (PDB code 1RFJ) with an rmsd of only 0.31 Å. Minor differences were observed at the C-terminal end, where Ca²⁺-binding loop region of potato CaM forms a small antiparallel β -sheet, whereas AtCaM7 forms loop (Fig. 2A). Structural superposition with other CaM, Animal CaM (PDB code 1UP5), *C. elegans* (PDB code 1OOJ), and *P. tetraurelia* (PDB code 1EXR) [32–34] suggest that AtCaM7 structure deviates from these mentioned CaM, as its rmsd value indicates (Table 2). Maximum deviation was found with *P. tetraurelia* CaM followed by animal CaM and *C. elegans*. These structural differences occur due to difference in overall length (extended conformation length) of the CaM and interhelical angles of each EF-hand motifs (Fig. 2B), although overall fold remains same. Interhelical angle comparison suggests that AtCaM7 EF-2 has more open hydrophobic pocket with respect to other CaM's EF-2. The rest of three EF-hand motifs of AtCaM7 have almost similar open hydrophobic pockets in comparison with other CaM's EF-hand motifs. Differences in the interhelical angles of various CaM's EF-hand motifs are shown in Table 3. Solvent assessable area

was calculated using the PISA server [35] and it was found that AtCaM7 structure has the highest solvent assessable area of 9868.5 Å² in comparison with other CaM, mentioned above. Recently it has been shown that MPD molecule is likely to favor more open conformations of the EF-hands in the crystal of CaM [36]. This could be the reason of having high solvent-assessable area in case of AtCaM7.

Arginine 127th of AtCaM7 is crucial residue for interaction with Z-box (AtCaM7 interacts with Z-box)

It was reported earlier that AtCaM7 interacts with Z-box of LRE and its other isoform CaM2/3/5 fails to show a similar result, although CaM2/3/5 is different from AtCaM7 by single a residue at position 127th (AtCaM7 has arginine residue at 127th position, whereas AtCaM2/3/5 has lysine) [17]. To further validate the interaction of AtCaM7 with Z-box DNA by biophysical method, BLI experiment was carried out using Z-box DNA with AtCaM7 and its isoform AtCaM2/3/5. The result clearly indicate that AtCaM2/3/5 did not interact with Z-box, as AtCaM2/3/5 sensor tip did not show change in signal response, suggesting that Z-box does not associate at AtCaM2/3/5 immobilized sensor tip, hence not showing interaction (Fig. 3A). However, AtCaM7 does interact with Z-box of LRE and signal from CaM7 sensor tip reaches up to 0.13 nm at 500 nM concentration of Z-box with K_D of 2.7 nM (Fig. 3B). Consistent with earlier reported result of AtCaM7-Z-box DNA binding [17], our BLI experiment also shows the same result as AtCaM7 found to interact with Z-box DNA and its isoforms AtCaM2/3/5 fail to interact. This result clearly suggests that Arg127th is the crucial residue responsible for the interaction between AtCaM7 and Z-box DNA.

Conformational sampling of AtCaM7 and AtCaM2/3/5

To explore the role of residue Arg127 in binding DNA, we modeled the AtCaM2/3/5 isoform from the crystal structure of AtCaM7 by using ROSETTA BACKRUB software [25]. Moreover, since CaM can adopt multiple conformations and since this flexibility and configurationally variability (i.e., the ability to alter the orientation and distance between its two domains) is also responsible for its diverse target recognition [37], we sought to determine whether the ability of AtCaM7 to bind DNA is dependent on the conformational changes or is just due to the single residue change. For this purpose, and to gain insight into the

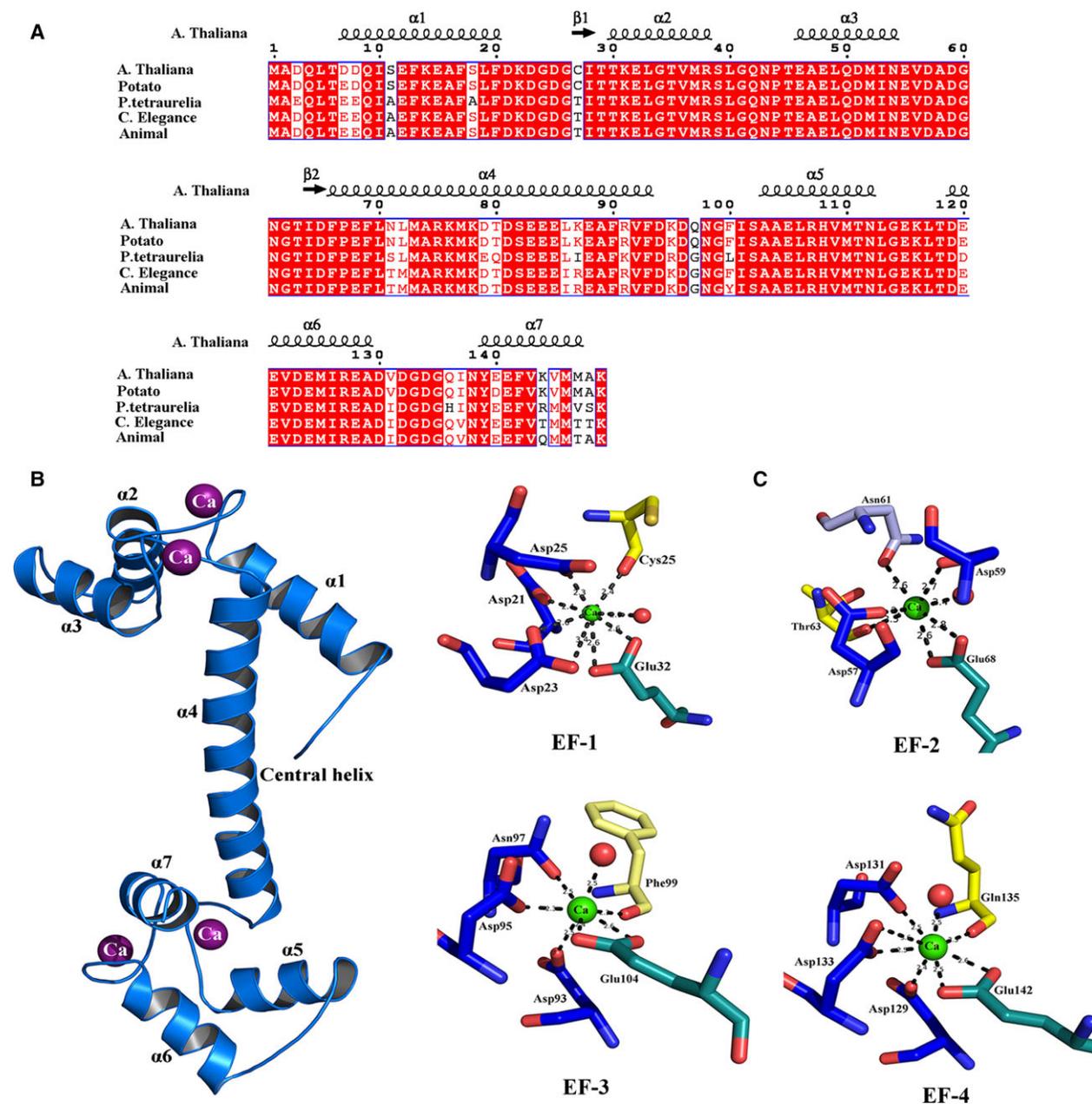


Fig. 1. Structural characterization of AtCaM7 (A) Sequence alignment of AtCaM7 with various CaM (Potato CaM, *P. tetraurelia*, *C. elegans*, and Animal CaM). (B) Overall structure of calmodulin 7 from *Arabidopsis thaliana*. In the crystal, AtCaM7 forms a dumbbell-shaped structure, consisting of two globular domains connected by a long central linker helix, similar to that seen in other calmodulin structures. Each domain contains two EF-hand motifs, and binds two calcium ions. (C) Each one of the four calcium-binding (calcium is shown in green) sites forms a pentagonal bipyramidal structure, where six calcium-coordinating oxygen atoms are from the protein and the 7th one is from water is shown in red.

conformational changes, we performed two independent Graphical Processing Unit (GPU)-based atomistic molecular dynamics simulations of AtCaM7 and AtCaM2/3/5 isoform for 150 ns on each system and captured snapshots at every 1 ps. The trajectories of both of these MD simulations stabilized at higher

rmsds, because of the formation of a hinge in the central helix that caused both molecules to form a relatively compact structure. The α -carbon rmsd calculated between the simulated structures and the crystal (or crystal-derived) structures showed similar trends for both AtCaM7 and AtCaM2/3/5. Besides

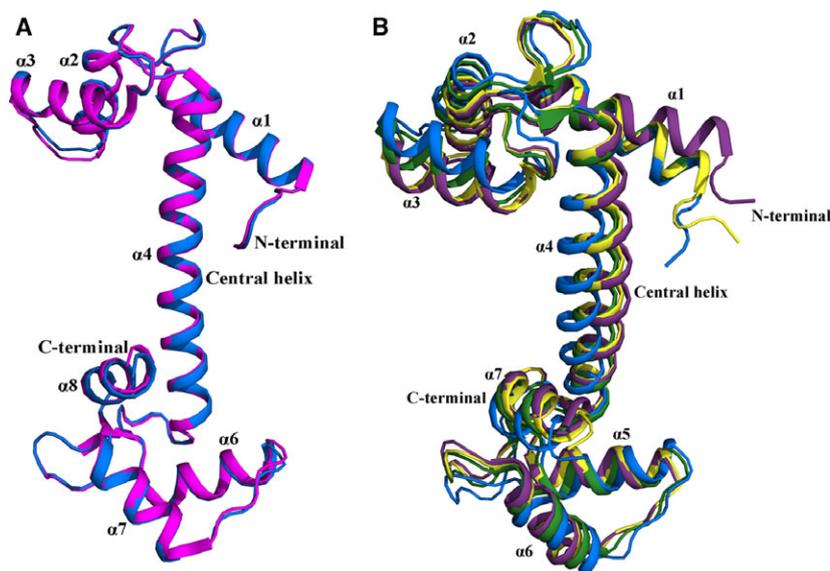


Fig. 2. Structural comparison (A) Structural comparison of AtCaM7 with potato calmodulin (pink, 1RFJ), indicating their very similar structures. (B) Structure alignment of AtCaM7 with calmodulin from *C. elegans* (PDB code 100J, green color) Animal (PDB code 1UP5, yellow color) and *P. tetraurelia* (PDB code 1EXR, violet color), represent structural differences.

Table 2. rmsd of Structural alignment of AtCaM7 with Calmodulin from other organism.

Organism	PDB code	rmsd (Å)
Potato CaM	1RFJ	0.31
<i>C. elegans</i>	100J	1.48
Chicken	1UP5	2.54
<i>P. tetraurelia</i>	1EXR	3.06

Table 3. Comparison of interhelical angles of EF-hand motifs of AtCaM7 with other CaM's EF-hand motifs.

Organism	PDB code	EF-1	EF-2	EF-3	EF-4
<i>A. thaliana</i>	5A2H	46.2	60.1	45.6	48.0
<i>P. tetraurelia</i>	1EXR	45.2	54.3	47.9	50.4
<i>C. elegans</i>	100J	45.1	54.4	44.3	49.3
Chicken	1UP5	44.4	52.1	48.6	51.0

differences during the multiple run, both the structures converged and showed similar characteristics in terms of the dynamics and folding (Fig. 4A,B). The residue wise root mean square fluctuations analysis (RMSFs) (Fig. 4C) showed that the major differences in the fluctuations occurred in the N-terminal domain, whereas fluctuations were almost similar in the C-terminal domain. Moreover the RMSF is almost same at position 127 suggesting that the movement or the structural changes in AtCaM7 is predominant in the N terminal. Furthermore, to understand the DNA binding, we calculated the solvent-accessible surface area (SASA) from all the snapshots taken from the MD trajectories of AtCaM7 and AtCaM2/3/5. The calculated SASA values of AtCaM7 and AtCaM2/3/5 from

the last 50 ns were similar, at 9622.79 and 9861.74 Å² (Fig. 4D). The overall distribution of available accessible surface suggested that the AtCaM7 is slightly more accessible than the isoform AtCaM2/3/5. Based on these MD simulations, we can suggest that major conformational changes do not play a role in driving the DNA binding. Hereby, suggesting that the difference in the side chain at position 127 is the main determinant of AtCaM7 to binds DNA.

AtCaM7-Z-box DNA complex model

It has been shown that AtCaM7 and Z-box DNA interact with each other and its isoform AtCaM2/3/5 fails to show same characteristics [17]. Molecular Dynamics simulation study of AtCaM7 and AtCaM2/3/5 suggests that both structures maintain a similar conformation upon stabilization after simulations. The initial model or the starting structure of the complex was obtained using the HADDOCK server [29]. During the docking simulations Arg127 was considered as the active site residue. Based on the cluster analysis with the highest score a final pose was obtained and was further refinement using the FIREDOCK module. The receptor or the AtCaM7 was docked in the major groove of the Z-box. Interestingly, the final docking pose of AtCaM7-DNA complex and this mode of binding resemble that of the HTH motif-DNA complex structures [38]. The model suggests that AtCaM7 and the Z-box DNA interact with each other in a perpendicular manner. Helix7, which is the part of EF-4 in AtCaM7 (contains the critical Arg127 residue) accommodated in the major groove of Z-box (Fig. 5A). Arg127 may form hydrogen

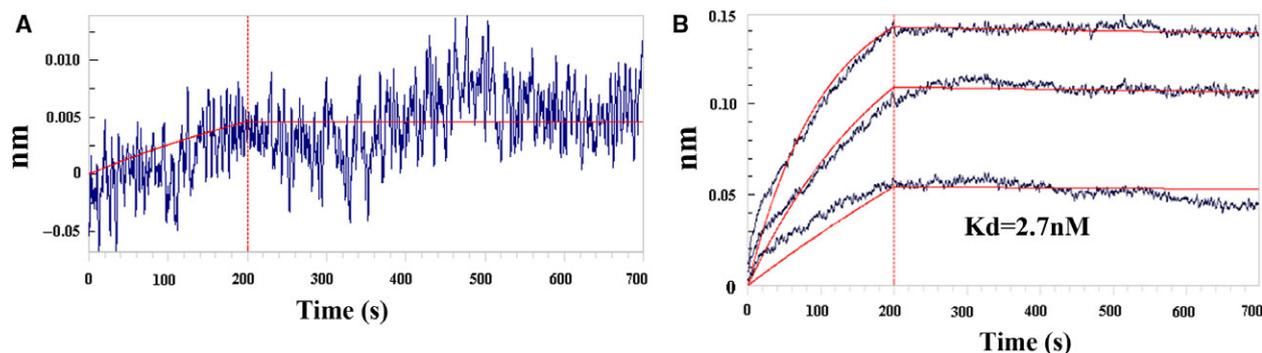


Fig. 3. Bi-layer interferometry (A) Binding Curve of CaM2/3/5 with Z-box DNA represent DNA not binding with AtCaM2/3/5-immobilized sensor chip as there is no change in interferometry pattern of AtCaM2/3/5-immobilized sensor tips at 500 nM concentration DNA. (B) Binding Curve of AtCaM7 with Z-box DNA represents CaM7–Z-box interaction, interferometry shift reaches up to 0.15 nm at 500 nM concentration of DNA.

bonds with either guanine 167 or 169, as these two guanines are in close proximity to Arg127. Interaction of arginine residue with guanine base is well supported by several transcriptional factor-DNA complexes, which shows that arginine has the highest expected ratio to interact with guanine [39]. Aside from Arg127, other residues that are expected to contribute in the binding of the Z-box are Met125 and Ile126 that could form hydrophobic interactions with a thymine base of 168 or 170.

Based on the docking scores and the interactions observed in the AtCaM7-Z box model, one can suggest that AtCaM7 binds Z-box via Arg-127. This side-ways interaction as seen with the best binding pose may be required for AtCaM7 to be able to dissociate readily from the Z-box so that it can also fulfill its other downstream signaling roles. In the case of the AtCaM2/3/5, the docking score clearly suggests that it likely does not bind with Z-box (Table 4). But to prove it furthermore and to understand the structural stability and binding affinity of the bound models of DNA with AtCaM7 and AtCaM2/3/5 (which could shed light to understand the structural basis of the impairment of AtCaM2/3/5 binding), we computed the stability (ΔG) of the complex using the empirical effective energy function of FOLDX [40]. FOLDX was used to calculate the changes in the binding affinity of the Z-box to AtCaM7 and AtCaM2/3/5. From this model, the energies associated with all interactions that were formed or lost upon mutation were quantified and used to calculate the predicted change in the native thermodynamic stability ($\Delta\Delta G_{\text{total}} = \Delta G_{\text{AtCaM2/3/5}} - \Delta G_{\text{AtCaM7}}$). We found an overall positive change in the $\Delta\Delta G$ for AtCaM2/3/5, indicating that AtCaM2/3/5 is less stable along with Z-box (or if when complex formed) and hence not able to hold

Z-box DNA. The overall change in terms of stability of the (CaM7-Z box) complex when mutated to AtCaM2/3/5 was $0.18 \text{ kcal}\cdot\text{mol}^{-1}$ and the $\Delta\Delta G$ was calculated as $0.414 \text{ kcal}\cdot\text{mol}^{-1}$.

Discussion

We have determined the crystal structure of AtCaM7 to 2.2 Å resolution. Structural analysis indicates that AtCaM7 structure is very similar to other CaM structures, which have been reported so far. Structural comparison with various CaM structure suggests that AtCaM7 shows minor differences in overall structure although overall fold remains the same. Bio layer interferometry study confirms that AtCaM7 interacts with Z-box DNA and Arg127th is the crucial residue for this interaction. While there are several reports indicating that Lys can also interact with DNA [41], Lys127-substituted isoforms AtCaM2/3/5 of this case could not interact with Z-box. The molecular dynamics simulations suggest that, compared to Lys127, Arg127 can adopt a more favorable, extended conformation, and be able to ‘hold onto’ the DNA. The extended Arg127 side chain can interact relatively closely with guanine of the Z-box, while the lysine either cannot reach so far or adopts conformations that place it away from the guanine (Fig. 5B). Due to these structural differences between the arginine and lysine, we expect that this favorable state of Arg127 is responsible for the AtCaM7-Z-box association. We also expect Arg127 to interact with a thymine base, which is located next to the guanine base with stacking and hydrophobic interactions. The arginine residue, with its guanidinium group, is superior to lysine for interaction with the guanine base: the guanidinium group has a planar structure, which can also stack on other

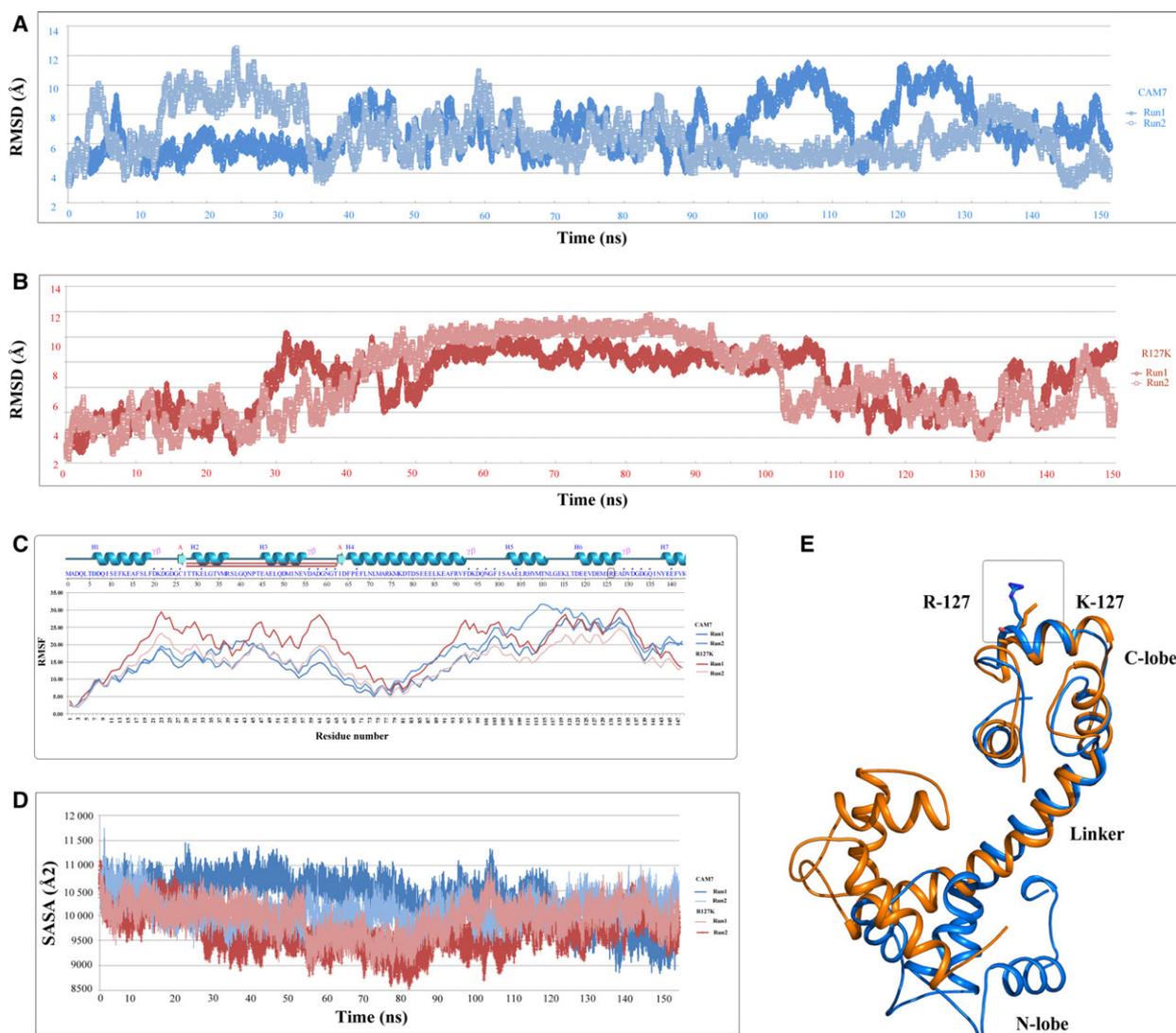


Fig. 4. Molecular dynamics simulations of the AtCaM7 structure and of the CaM-AtCaM2/3/5 model. (A) Protein backbone rmsds of the AtCaM7 crystal structure from two simulation runs of 150 ns each. (B) Protein backbone rmsds between the AtCaM2/3/5 model to two simulation runs of 150 ns each. (C) Root mean square fluctuations from the multiple molecular dynamics (MMD) simulation run of AtCaM7 and AtCaM2/3/5 for every residue. The secondary structures as well as the sequence of AtCaM7 are shown at the top of the plot. (D) Computed solvent-accessible surface area (SASA) of AtCaM7 and AtCaM2/3/5 from the MMD simulation runs. (E) Superposition of average structure of AtCaM7 and AtCaM2/3/5 extracted from the MMD Simulation trajectory.

bases, whereas lysine has only one amino group, allowing interaction only in one direction [42]. This specific property of arginine allows it to form more hydrogen bonds, salt bridges, and other electrostatic interactions than lysine can. In addition to this geometric effect, the ionic interactions formed by arginine are often more stable than those that are formed by lysine, particularly under alkaline pH, due to the higher pKa of arginine than of lysine [42]. We expect that such advantages of arginine relative to lysine, causes AtCaM7 to be able to interact with the Z-box DNA.

The energetically minimized complex model of AtCaM7-Z-box indicates a possible conformation/mode of interaction between AtCaM7 and the Z-box DNA. In which they are oriented perpendicularly to one another and AtCaM7 adopt extended conformation. After simulations the CaM7 adopts bent structure where N- and C-terminal domains move closer. Even this conformation can also bind to DNA with C-terminal domain, as there were no changes seen in C-terminal domain in simulations. The Ca^{2+} -defective mutants of AtCaM7 fail to bind with DNA, even if

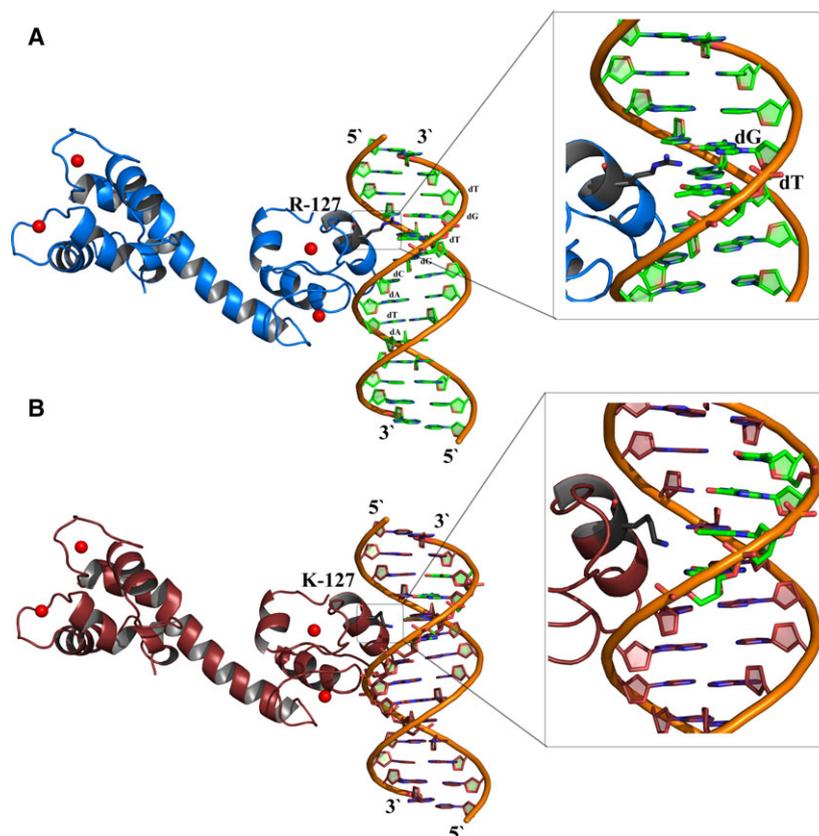


Fig. 5. Molecular docking simulations of AtCaM7 and AtCaM2/3/5 with DNA. (A) The model was obtained after performing HADDOCK-based docking simulations on AtCaM7 with DNA-containing Z-box. The zoomed inset clearly shows Arg127 interacting with a guanine of the DNA. (B) Complex Model of AtCaM2/3/5 with DNA.

Table 4. Docking scores obtained from HADDOCK server for AtCaM7 – Z-box DNA and AtCaM2/3/5 – Z-box DNA.

Docking parameters	AtCaM7	AtCaM2/3/5
HADDOCK score (kcal·mol ⁻¹)	-15.5 ± 13.7	78.9 ± 2.5
Cluster size (no. of clusters)	10	10
rmsd from the overall lowest-energy structure (Å)	0.3 ± 0.2	0.3 ± 0.2
Van der Waals energy (kcal·mol ⁻¹)	-21.1 ± 1.9	-10.3 ± 1.2
Electrostatic energy (kcal·mol ⁻¹)	-389.7 ± 26.4	-2.9 ± 7.0
Desolvation energy (kcal·mol ⁻¹)	4.1 ± 5.5	16.7 ± 2.1
Restraints violation energy (kcal·mol ⁻¹)	95.1 ± 26.03	730.7 ± 3.15
Buried surface area (Å ²)	319.7 ± 23.7	329.7 ± 27.7

only EF-3 was mutated and also similar results were obtained when EF-1 and EF-3, both were mutated [17]. We therefore expect that removal of Ca²⁺ from EF3 or EF-1 would disrupt the structural integrity of the CaM, hence DNA binding [17]. Moreover, previous molecular dynamics simulation indicated that removal of even a single Ca²⁺ ion could induce unfolding of CaM, and hence the overall structure adopts a closed conformation [2]. Therefore, we expect that mutation of even a single EF-hand motif is sufficient to change the conformation of CaM, and when done in the C-terminal domain would abolish structural integrity of the C-terminal domain and thus the

DNA interactions [17]. There are few structures reported (PDB code 4DCK, 4DJC, 4E53, and 4E50) where CaM is observed in extended conformation after binding with its target molecules [11,43,44]. These earlier reported studies suggest our model that AtCaM7 possibly adopts extended conformation for binding with Z-box DNA or AtCaM7 can also form compact structure after binding to DNA. Based on the BLI interaction data, molecular dynamics, and docking study, we propose that R127 of AtCaM7 is a crucial residue for Z-box DNA binding, whereas K-127 of AtCaM2/3/5 lacks such a bidentate side chain and hence fails to bind the DNA.

Acknowledgements

We thank the Department of Biotechnology, Government of India for funding. SK, MM, thank DBT and ICMR for their fellowship. We would also like to thank Sushilendra Vaidya and Mohit Verma from Pal Life Science for bringing Octet K2 instrument for demonstration and helping in protein–DNA interaction work. We thank ESRF, BM 14 staff, Dept. of Biotechnology (DBT), Government of India for access to the beam line and National Institute of Immunology for data collection. We thank UGC-RNW, BDT-BUILDER, UGC-UPE2, DST-PURSE, DST-FIST for institutional and central instrumentation facility support.

Author contributions

SK, MM, SC, and SG designed the experiment, wrote the manuscript, analyzed the data, and drew pictures. SK and NG purified the protein; SK and SG determined the structure; MM performed molecular simulations and docking. All authors reviewed the manuscript.

References

- Park HY, Kim SA, Korlach J, Rhoades E, Kwok LW, Zipfel WR, Waxham MN, Webb WW and Pollack L (2008) Conformational changes of calmodulin upon Ca²⁺ binding studied with a microfluidic mixer. *Proc Natl Acad Sci USA* **105**, 542–547.
- Zhang Y and Lou J (2012) The Ca(2+) influence on calmodulin unfolding pathway: a steered molecular dynamics simulation study. *PLoS One* **7**, e49013.
- Lewit-Bentley A and Rety S (2000) EF-hand calcium-binding proteins. *Curr Opin Struct Biol* **10**, 637–643.
- Gifford JL, Walsh MP and Vogel HJ (2007) Structures and metal-ion-binding properties of the Ca²⁺-binding helix-loop-helix EF-hand motifs. *Biochem J* **405**, 199–221.
- Ikura M (1996) Calcium binding and conformational response in EF-hand proteins. *Trends Biochem Sci* **21**, 14–17.
- Vetter SW and Leclerc E (2003) Novel aspects of calmodulin target recognition and activation. *Eur J Biochem* **270**, 404–414.
- Shen X, Valencia CA, Szostak JW, Dong B and Liu R (2005) Scanning the human proteome for calmodulin-binding proteins. *Proc Natl Acad Sci USA* **102**, 5969–5974.
- Tidow H, Poulsen LR, Andreeva A, Knudsen M, Hein KL, Wiuf C, Palmgren MG and Nissen P (2012) A bimodular mechanism of calcium control in eukaryotes. *Nature* **491**, 468–472.
- Zhang M, Abrams C, Wang L, Gizzi A, He L, Lin R, Chen Y, Loll PJ, Pascal JM and Zhang JF (2012) Structural basis for calmodulin as a dynamic calcium sensor. *Structure* **20**, 911–923.
- Chin D and Means AR (2000) Calmodulin: a prototypical calcium sensor. *Trends Cell Biol* **10**, 322–328.
- Wang C, Chung BC, Yan H, Lee SY and Pitt GS (2012) Crystal structure of the ternary complex of a NaV C-terminal domain, a fibroblast growth factor homologous factor, and calmodulin. *Structure* **20**, 1167–1176.
- Houdusse A, Gaucher JF, Krementsova E, Mui S, Trybus KM and Cohen C (2006) Crystal structure of apo-calmodulin bound to the first two IQ motifs of myosin V reveals essential recognition features. *Proc Natl Acad Sci USA* **103**, 19326–19331.
- Corneliussen B, Holm M, Waltersson Y, Onions J, Hallberg B, Thornell A and Grundstrom T (1994) Calcium/calmodulin inhibition of basic-helix-loop-helix transcription factor domains. *Nature* **368**, 760–764.
- Craig TA, Benson LM, Venyaminov SY, Klimtchuk ES, Bajzer Z, Prendergast FG, Naylor S and Kumar R (2002) The metal-binding properties of DREAM: evidence for calcium-mediated changes in DREAM structure. *J Biol Chem* **277**, 10955–10966.
- Carrion AM, Link WA, Ledo F, Mellstrom B and Naranjo JR (1999) DREAM is a Ca²⁺-regulated transcriptional repressor. *Nature* **398**, 80–84.
- Gilchrist CA, Holm CF, Hughes MA, Schaenman JM, Mann BJ and Petri WA Jr (2001) Identification and characterization of an *Entamoeba histolytica* upstream regulatory element 3 sequence-specific DNA-binding protein containing EF-hand motifs. *J Biol Chem* **276**, 11838–11843.
- Kushwaha R, Singh A and Chattopadhyay S (2008) Calmodulin7 plays an important role as transcriptional regulator in *Arabidopsis* seedling development. *Plant Cell* **20**, 1747–1759.
- Abbas N, Maurya JP, Senapati D, Gangappa SN and Chattopadhyay S (2014) *Arabidopsis* CAM7 and HY5 physically interact and directly bind to the HY5 promoter to regulate its expression and thereby promote photomorphogenesis. *Plant Cell* **26**, 1036–1052.
- Otwinowski Z and Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307–326.
- Vagin A and Teplyakov A (2010) Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66**, 22–25.
- Yun CH, Bai J, Sun DY, Cui DF, Chang WR and Liang DC (2004) Structure of potato calmodulin PCM6: the first report of the three-dimensional

- structure of a plant calmodulin. *Acta Crystallogr D Biol Crystallogr* **60**, 1214–1219.
- 22 Emsley P and Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132.
- 23 Murshudov GN, Vagin AA and Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240–255.
- 24 Laskowski RA, MacArthur MW, Moss DS and Thornton JM (1993) Procheck – a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26**, 283–291.
- 25 Lauck F, Smith CA, Friedland GF, Humphris EL and Kortemme T (2010) RosettaBackrub – a web server for flexible backbone protein structure modeling and design. *Nucleic Acids Res* **38**, W569–W575.
- 26 Hornak V, Abel R, Okur A, Strockbine B, Roitberg A and Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–725.
- 27 DeLano WL and Lam JW (2005) PyMOL: a communications tool for computational models. *Abstr Pap Am Chem Soc* **230**, U1371–U1372.
- 28 van Dijk M and Bonvin AM (2009) 3D-DART: a DNA structure modelling server. *Nucleic Acids Res* **37**, W235–W239.
- 29 de Vries SJ, van Dijk M and Bonvin AM (2010) The HADDOCK web server for data-driven biomolecular docking. *Nat Protoc* **5**, 883–897.
- 30 Mashiah E, Schneidman-Duhovny D, Andrusier N, Nussinov R and Wolfson HJ (2008) FireDock: a web server for fast interaction refinement in molecular docking. *Nucleic Acids Res* **36**, W229–W232.
- 31 Van Durme J, Delgado J, Stricher F, Serrano L, Schymkowitz J and Rousseau F (2011) A graphical interface for the FoldX forcefield. *Bioinformatics* **27**, 1711–1712.
- 32 Rupp B, Marshak DR and Parkin S (1996) Crystallization and preliminary X-ray analysis of two new crystal forms of calmodulin. *Acta Crystallogr D Biol Crystallogr* **52**, 411–413.
- 33 Symersky J, Lin G, Li S, Qiu S, Carson M, Schormann N and Luo M (2003) Structural genomics of *Caenorhabditis elegans*: crystal structure of calmodulin. *Proteins* **53**, 947–949.
- 34 Wilson MA and Brunger AT (2000) The 1.0 Å crystal structure of Ca(2+)-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. *J Mol Biol* **301**, 1237–1256.
- 35 Krissinel E and Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774–797.
- 36 Lin J, van den Bedem H, Brunger AT and Wilson MA (2016) Atomic resolution experimental phase information reveals extensive disorder and bound 2-methyl-2,4-pentanediol in Ca(2+)-calmodulin. *Acta Crystallogr D Struct Biol* **72**, 83–92.
- 37 Tidow H and Nissen P (2013) Structural diversity of calmodulin binding to its target sites. *FEBS J* **280**, 5551–5565.
- 38 Gao X, Zou T, Mu Z, Qin B, Yang J, Waltersperger S, Wang M, Cui S and Jin Q (2013) Structural insights into VirB-DNA complexes reveal mechanism of transcriptional activation of virulence genes. *Nucleic Acids Res* **41**, 10529–10541.
- 39 Luscombe NM, Laskowski RA and Thornton JM (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res* **29**, 2860–2874.
- 40 Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F and Serrano L (2005) The FoldX web server: an online force field. *Nucleic Acids Res* **33**, W382–W388.
- 41 Azorin F, Vives J, Campos JL, Jordan A, Lloveras J, Puigjaner L, Subirana JA, Mayer R and Brack A (1985) Interaction of DNA with lysine-rich polypeptides and proteins. The influence of polypeptide composition and secondary structure. *J Mol Biol* **185**, 371–387.
- 42 Sokalingam S, Raghunathan G, Soundararajan N and Lee SG (2012) A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS One* **7**, e40410.
- 43 Sarhan MF, Tung CC, Van Petegem F and Ahern CA (2012) Crystallographic basis for calcium regulation of sodium channels. *Proc Natl Acad Sci USA* **109**, 3558–3563.
- 44 Kumar V, Chichili VP, Zhong L, Tang X, Velazquez-Campoy A, Sheu FS, Seetharaman J, Gerges NZ and Sivaraman J (2013) Structural basis for the interaction of unstructured neuron specific substrates neuromodulin and neurogranin with Calmodulin. *Sci Rep* **3**, 1392.