Journal of Biomolecular NMR **30:** 451–456, 2004.

Letter to the Editor: Solution structure of a calmodulin-like calcium-binding domain from *Arabidopsis thaliana**

Jikui Song, Qin Zhao, Sandy Thao, Ronnie O. Frederick & John L. Markley**
Center for Eukaryotic Structural Genomics, Department of Biochemistry, University of Wisconsin-Madison, WI 53706-1544, USA

Received 8 July 2004; Accepted 30 September 2004

Key words: calmodulin, EF-hand, structural genomics

Biological Context

Calmodulin (CaM) is a ubiquitous Ca^{2+} binding protein highly conserved in all eukaryotes. By responding to intracellular Ca^{2+} signaling and by binding to a large array of different protein targets, the protein regulates numerous cellular signal transduction events. At present, information about the structural and functional properties of CaMs and CaM like proteins (CML) derives primarily from studies of proteins from animals; thus the physiological roles of CaMs in plants are relatively unclear. Nevertheless, in recent years the sequences of a growing number of plant CaMs and CaM related proteins have been identified: for example, 6 CaMs and 50 additional CMLs from *Arabidopsis thaliana* alone. These *Arabidopsis* CaMs share a 91% sequence identity with those from vertebrates, while sequence identities between the CMLs and the CAMs within the same *Arabidopsis* species vary from 16.1 to 73.2% (McCormack and Braam, 2003).

An open reading frame (ORF) from *Arabidopsis thaliana* (*At*3g03410.1) coding for a CML was chosen for structure determination by the Center for Eukaryotic Structural Genomics (CESG). This CML, which consists of 131 amino acids, was shown to have 35.9% sequence identity with the known *Arabidopsis* CaM, CaM2 (Figure 1). Although the N-terminus of *At*3g03410.1 is 8 residues shorter than that of CaM2, the protein is still predicted to conform to the CaM paradigm of four EF-hands, with each of the four Ca^{2+}-binding pockets consisting of 12 contiguous residues containing an invariant Asp and Glu at the 1st and the 12th positions, respectively. An initial NMR investigation of the full-length *At*3g03410.1 gene product, however, showed it to be an intractable target for NMR studies. To simplify the system, we selected for structure determination the Ca^{2+}-loaded N-terminal domain of *At*3g03410.1 (Ca-nCML), which has 31% sequence identity with the corresponding part from CaM (Figure 1).

Materials and methods

The *At*3g03410.1 ORF was amplified directly from *Arabidopsis thaliana* genomic DNA. A two-step PCR protocol was used to add the Gateway® (Invitrogen Inc.) recombination sites and a tobacco etch virus (TEV) protease cleavage site to the ORF (Aceti et al., 2003). The gene was transferred to pDEST17 (Invitrogen Inc.) for protein expression, and the DNA sequence was verified. Site-directed mutagenesis was used to reduce the size of the coding region to one corresponding to the N-terminal domain of *At*3g03410.1 (residue 1–66) (nCML). The construct was transformed into Rosetta(DE3)/pLysS (Novagen) host cells, and these were used according to an established protocol (Zhao et al., 2004)

*Structure data have been deposited at PDB (1TIZ) and NMR data at BMRB (bmr 6209).

**To whom correspondence should be addressed. E-mail: markley@nmrfam.wisc.edu

Figure 1. Alignment of the sequences of *At*3g03410.1, CaM2 and nCML from *Arabidopsis thaliana* with conserved regions highlighted in green. Residues marked by "⋆" indicate the hypothetical Ca^{2+} binding sites in *At*3g03410.1.

in preparing samples labeled with ^{15}N or $^{15}N/^{13}C$. The cells were harvested by centrifugation and stored overnight at −20 °C. The cell pellet was suspended in 50 ml lysis buffer (50 mM Tris, 100 mM NaCl, pH 8.0) and sonicated on ice until a very low viscosity of the suspension was achieved. The pellet containing inclusion bodies was separated by centrifugation and washed with the lysis buffer in the presence of 1% (v/v) Triton X-100. The buffer was then removed by centrifugation, and the inclusion bodies were dissolved in loading buffer: 6 M guanidinium chloride (GdmCl), 20 mM sodium phosphate (NaP_i), 5 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 5 mM imidazole, pH 7.4. After removing the insoluble cell debris, the supernatant was loaded onto a precharged nickel nitrilotriaceitc acid (Ni-NTA) affinity column (QIAGEN) with equilibrated resin. Following extensive washes with the loading buffer, the protein was eluted with denaturing elution buffer (6 M GdmCl, 20 mM NaP_i, 250 mM imidazole, pH 7.4) and concentrated with an Amicon ultrafiltration device (Amersham Pharmacia, Inc.). The protein was refolded by the rapid addition of refolding buffer (20 mM Bis–Tris, 5 mM DTT, 5 mM Ca^{2+}, 0.2 M arginine HCl, pH 6.4). After refolding, ultrafiltration was used to concentrate the protein and to exchange it into the TEV reaction buffer (20 mM Tris, 1 mM DTT, 5 mM $CaCl_2$, pH 7.8). The protein was cleaved by incubation overnight with $(His)_6$-tagged TEV protease (prepared in house) (1/100, w/w) at 25 °C. After cleavage, the protein solution was again applied to a Ni-NTA column. The target protein was collected from the column flow-through, and ultrafiltration was used to concentrate the protein and to exchange it into the NMR buffer (90% H_2O/10% D_2O, 5 mM $CaCl_2$, 5 mM DTT, 20 mM Bis–Tris, and 0.02% (w/v) sodium azide, pH 6.4). The 67-residue protein product (Ca-nCML) included a non-native N-terminal serine left after TEV cleavage. Sample purity (>95%) and molecular weight were verified by SDS-PAGE and MALDI-TOF mass spectrometry.

Normal NMR samples contained 1 mM [U-^{15}N]- or [U-^{13}C,^{15}N]-Ca-nCML in the NMR buffer plus 1 mM DSS added as the internal chemical shift reference. The sample used for measurements of residual dipolar couplings contained 0.5 mM [U-^{15}N]-Ca-nCML in NMR buffer containing 5% pentaethylene glycol monododecyl ether.

All NMR spectra were recorded at the National Magnetic Resonance Facility at Madison (NMRFAM) on a 600 MHz Bruker DMX spectrometer equipped with a triple-resonance CryoProbe. The temperature of each sample was held at 25 °C. The [U-^{15}N]-Ca-nCML sample was used in acquiring 3D ^{15}N-edited NOESY-HSQC (τ_{mix} = 125 ms) and TOCSY-HSQC data sets. The [U-^{13}C,^{15}N]-Ca-nCML sample was used in acquiring 3D HNCACB, HN(CO)CACB, HNCO, HCCONH, CCONH, HCCH-COSY, ^{13}C edited [^1H, ^1H] NOESY (τ_{mix} = 125 ms) data sets. A 2D IPAP ^1H-^{15}N HSQC experiment (Ottiger et al., 1998) was used to obtain the one-bond N–H scalar coupling ($^1J_{N-H}$) values. Residual dipolar couplings for N–H vectors ($^1D_{N-H}$) were calculated from differences in scalar couplings ($^1J_{N-H}$) measured in the presence and absence of the ordering agent (5% pentaethylene glycol monododecyl ether).

Spectra were processed and analyzed, respectively, with the NMRPipe (Delaglio et al., 1995) and Sparky (http://www.cgl.ucsf.edu/home/sparky) software packages. The GARANT software package was used in assigning the backbone signals. Assignments of proton resonances from aromatic

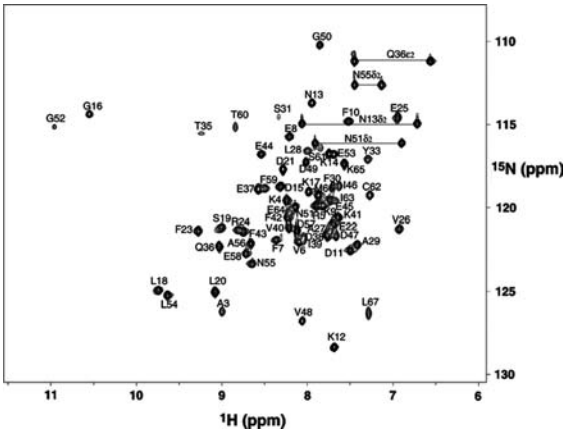

Figure 2. [^{1}H, ^{15}N]-HSQC spectrum of the Ca-nCML from *Arabidopsis thaliana* recorded at 600 MHz ^{1}H resonance frequency. The protein concentration was ~1 mM, the Ca^{2+} concentration was 5 mM, the pH was 6.4, and the temperature was 25 °C.

side chains or side chain amide groups of Asn and Gln were based on NOEs between these protons and the $^1H^{\beta 2\beta 3}$ protons, as observed in 3D ^{13}C-edited NOESY-HSQC and 3D ^{15}N-edited NOESY-HSQC spectra. The backbone assignments were 99% complete (Figure 2), and the sidechain assignments were ~90% complete. The time-domain NMR data and chemical shift assignments have been deposited in the BioMagResBank database under BMRB accession number 6209.

TALOS (Cornilescu et al., 1999) and CSI (Wishart and Sykes, 1994) programs were used to predict dihedral angles ϕ and ψ and hydrogen bond constraints. The automated CANDID iterative refinement module of Cyana (Herrmann et al., 2002) was used in generating initial NOE assignments and an initial set of structural models. Additional NOE assignments were then added and erroneous ones were corrected through examination of NMR spectra prior to recalculation of structures by Cyana (Güntert et al., 1997). The simulated annealing protocol in XPLOR-NIH (Schwieters et al., 2003) was used for the refinement that incorporated dipolar coupling restraints; this refinement was carried out to ensure the accurate determination of the relative orientations of individual structural elements. Prior to inclusion of the dipolar couplings, the program PALES (Zweckstetter and Bax, 2000) gave a correlation coefficient of $R_{svd} = 0.94$, corresponding to Q-factor of 33%. Following the incorporation of dipolar couplings, the R_{svd} and Q-factor improved to 0.98 and 18% respectively, although the overall structure did not change significantly. A total of 1510 NOE distance restraints, 100 backbone ϕ and ψ constraints, 46 hydrogen bond constraints, and 55 backbone NH dipolar coupling constraints were used as input for the final calculation of the ensemble of conformers representing the structure of Ca-nCML (Table 1). PROCHECK was used to analyze the quality of the structures, and MOLMOL (Koradi et al., 1996) was used to evaluate r.m.s.d. values. The three-dimensional coordinates for these models have been deposited in the Protein Data Bank (PDB) under the accession number 1TIZ.

Table 1. Statistics for the 20 energy-minimized conformers of Ca-nCML

Distance constraints	
Long [$(i-j)>5$]	276
Medium [$1<(i-j)\leq 5$]	482
Sequential [$(i-j)=1$]	378
Intraresidue [$i=j$]	374
Dihedral angle constraints (ϕ and ψ)	100
Hydrogen bond constraints	46
Dipolar coupling restraints (NH)	55
Average pairwise r.m.s.d. to the mean structure (Å)	
Residues 2–67	
Backbone (C^α, C′, N, O)	0.27 ± 0.09
Heavy atoms	0.87 ± 0.11
Regular secondary structural elements	
Backbone (C^α, C′, N, O)	0.21 ± 0.08
Heavy atoms	0.84 ± 0.09
Deviations from idealized covalent geometry	
Bonds (Å)	0.0068 ± 0.0001
Angles (°)	0.83 ± 0.02
Impropers (°)	0.75 ± 0.02
Ramachandran statistics (% of all residues)	
Most favored	87.5
Additionally allowed	9.3
Generously allowed	3.3
Disallowed	0.0

Results and discussion

Statistics on the structure are summarized in Table 1. Residue Ala29, which is adjacent to the

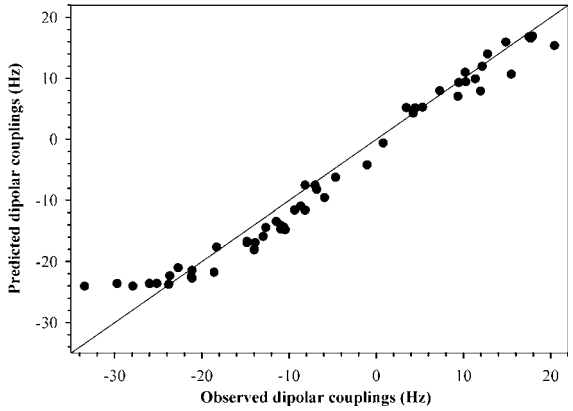

Figure 3. Correlation between the observed backbone NH dipolar couplings and couplings predicted by using an alignment tensor obtained from the SVD fit to the lowest-energy conformer of *Arabidopsis* Ca-nCML (from the set of 20 obtained with a set of restraints including RDCs).

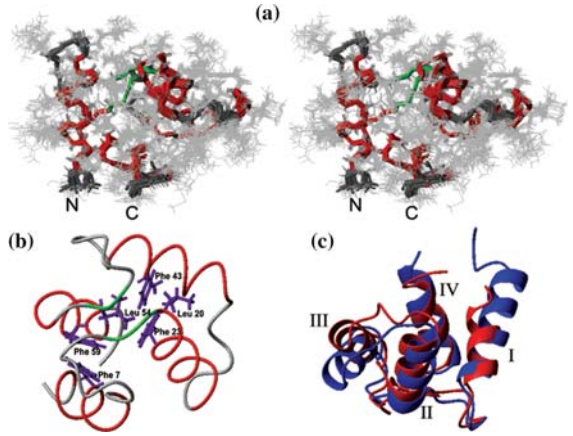

Figure 4. (a) Stereoscopic view of the final family of 20 conformers representing the structure of *Arabidopsis* Ca-nCML. The colors for backbone represents: (red) α-helix, (green) β-strand, and (dark grey) other regions. The side chains of the protein are shown in light gray. (b) Side chains of some of the hydrophobic residues (shown in purple) in Ca-nCML. The color scheme for backbone atom is the same as that for (a). (c) Ribbon diagram of the backbone of *Arabidopsis* Ca-nCML in red superimposed with the N-terminal domain of human Ca-CaM (PDB accession number, 1J7O) in blue. Helices are labeled with Roman numerals: I (α_1), II (α_2), III (α_3), IV (α_4).

linker between the two EF-hands, was one of the major outliers in the SVD fit (Losonczi et al., 1999) carried out on the structure refined with dipolar couplings (Figure 3). This may indicate mobility in this region. Figure 4a shows a stereo view of the family of 20 conformers representing the structure of Ca-nCML. The overall fold of Ca-nCML, which is very similar to those of each of the two domains of CaM, consists of two closely packed EF-hand motifs connected by a short hydrophobic turn. The first EF-hand is formed by helices α_1 (residues 2–11), α_2 (20–30) and the loop between them; and the second EF-hand is formed by helices α_3 (35–47), α_4 (55–65) and the interhelical loop. As anticipated from the sequence alignment (Figure 1), the length of helix α_1 in Ca-nCML is significantly shorter than that of α_1 in CaM (15–16 residues), whereas the lengths of the other three helices are similar in the two proteins. In addition to the four helices, a mini anti-parallel β-sheet is formed by strands β_1 (18–19) and β_2(53–54), as seen in CaM. The two Ca^{2+} binding sites, each of which consists of 12 consecutive residues in individual EF-hands, are mainly located in a loop between the two helices and at the beginning of the second helix. The structure contains many hydrophobic contacts. In particular, a hydrophobic cluster, formed by residues Leu20, Phe23, Phe43, and Leu54, helps hold the domain together. Two conserved phenylalanine residues (Phe7 and Phe59) are spatially close to one other (C^{α}–C^{α} distance = 7.3 Å) (Figure 4b). Interactions between two phenylalanines conserved at these positions are frequently observed in EF-hand containing proteins (Rashidi et al., 1999).

One of the prominent characteristics of CaM is a Ca^{2+}-induced conformational readjustment. Upon Ca^{2+} binding, the two helices of each EF-hand change from a 'closed' conformation to an 'open' conformation: this corresponds to a change of the interhelical angle within each EF-hand from ~130° to ~100° (Kuboniwa et al., 1995; Chou et al., 2001). The interhlx program (Yap, University of Toronto) was used to compute the interhelical angles in the individual EF-hands of the Ca-nCML. The results yielded average values (over the 20 conformers) of 102° ± 2 for the first EF-hand and 147° ± 2 for the second EF-hand. These values are compared in Table 2 with those from NMR structures of other domains: Ca-CaM, Ca^{2+}-free CaM, Ca^{2+}-ligated calbindin, and Ca^{2+}-free calbindin (Skelton et al., 1994; Kordel et al., 1997; Chou et al., 2000, 2001). The interhelical angle for the first EF-hand of Ca-nCML is equivalent within the experimental error to that of Ca-CaM. Surprisingly, the value for the second EF-hand in

Table 2. Comparison of the interhelical angles (between helices I and II, and III and IV) in Ca-nCML with those in the N-terminal domains of CaM and calbindin

	I—II	III–IV	PDB entry
Ca-nCML	102 ± 2	147 ± 3	1TIZ
Ca^{2+}-ligated CaM	104 ± 1	102 ± 2	1J7O
Ca^{2+}-free CaM	131 ± 1	130 ± 1	1F70
Ca^{2+}-ligated calbindin	126 ± 4	114 ± 6	1B1G
Ca^{2+}-free calbindin	119 ± 4	119 ± 8	1CLB

Ca-nCML differs significantly from that of Ca-CaM and is even higher than that of Ca^{2+}-free CaM. Despite this difference, the relative orientations of helices α_1 and α_4 in the two proteins are similar. To our knowledge, such a closed EF-hand has never been reported before in a Ca^{2+}-loaded CaM, although it has been observed in another class of EF-hand protein, calbindin D_{9k}. However, in contrast to Ca-nCML, both of the two EF-hands of Ca^{2+}-loaded calbindin D_{9k} were reported to be closed (Table 2) (Skelton et al., 1994).

The structural difference between Ca-nCML and the N-terminal domain of Ca-CaM is best visible by optimizing the superposition of helices α_1, α_2 and α_4 in these two structures. This leads to a close fit of all of the secondary structural elements (Figure 4c), with the exception of α_3, and alignment of the two Ca^{2+} binding sites with r.m.s.d. values of 0.63 and 0.72 Å, respectively. The most divergent regions of the two proteins include helix α_3 and the loop between the two EF-hands. Mutagenesis studies have shown that the composition of amino acids in the loop between the two EF-hands is critical in determining the conformation of the EF-hands and is partially responsible for the difference between CaM and calbindin D_{9k} in their response to Ca^{2+} binding (Nelson et al., 2002). The sequence identity between *At*3g03410.1 and CaM within this loop is low, and this stretch is more hydrophobic in *At*3g03410.1 than in CaM (Figure 1). For example, Pro43 in CaM is replaced by Phe34 in *At*3g03410.1. The fact that the backbone resonances from residues Phe30, Ser31, Tyr33 and Phe34 show considerable NMR signal broadening or heterogeneity suggests that this region may be dynamically disordered or involved in protein–protein interactions. Ca-nCML exhibited other signs of partial aggregation: earlier than expected elution from a gel exclusion column and broader than expected HSQC peaks (Figure 2). We used pulsed field gradient spectroscopy to measure the translational diffusion coefficient of Ca-nCML at different concentrations. These results (Supplementary Materials) indicated that Ca-nCML is only about 50% monomeric under the experimental conditions used in the structure determination.

Conclusion

In this study, the solution structure of the Ca^{2+}-loaded form of the N-terminal domain of a CML protein from *Arabidopsis thaliana* (Ca-nCML) has been determined by NMR spectroscopy. The domain contains two EF-hands packed closely with each other. The four α-helices and a small β-sheet are conserved between Ca-nCML and Ca-CaM, although the first α-helix is slightly shorter in Ca-nCML than in CaM. The interhelical angles of Ca-nCML indicate that the conformation of the first EF-hand is 'open', whereas that of the second EF-hand is 'closed'. Thus Ca-nCML differs from the N-terminal domain of Ca-CaM in which both hands are 'open'. Ca-nCML also differs from the Ca^{2+}-loaded form of calbindin in which both EF-hands are 'closed'. Although the underlying features are unclear at present, these differences are of interest. It has been suggested that the divergent conformational changes of CaM and calbindin in response to calcium binding may explain their different functional roles in the cell (Skelton et al., 1994).

Acknowledgements

This research was supported by the NIH Protein Structure Initiative through Grant 1 P50 GM64598. NMR data were collected and analyzed in the National Magnetic Resonance Facility at Madison, which is supported by National Institutes of Health Grants P41RR02301 (Biomedical Research Technology Program, National Center for Research Resources) and P41GM66326 (National Institute of General Medical Sciences). Equipment in the facility was purchased with

funds from the University of Wisconsin, the National Institutes of Health (P41GM66326, P41RR02301, RR02781, RR08438), the National Science Foundation (DMB-8415048, OIA-9977486, BIR-9214394), and the U.S. Department of Agriculture.

References

Aceti, D.J., Blommel, P.G., Endo, Y., Fox, B.G., Frederick, R.O., Hegeman, A.D., Jeon, W.B., Kimball, T.L., Lee, J.M.. Newman, C.S., Peterson, F.C., Sawasaki, T., Seder, K.D., Sussman, M.R., Ulrich, E.L., Wrobel, R.L., Thao, S., Vinarov, D.A., Volkman, B.F. and Zhao, Q. (2003) In *Biopolymers*. Steinbuchel, A. (ed.), Weinheim, Wiley-VCH, pp. 469–496.

Chou, J.J., Li, S. and Bax, A. (2000) *J. Biomol. NMR*, **18**, 217–227.

Chou, J.J., Li, S., Klee, C.B. and Bax, A. (2001) *Nat. Struct. Biol.*, **8**, 990–997.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR*, **13**, 289–302.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298.

Herrmann, T., Guntert, P. and Wüthrich, K. (2002) *J. Mol. Biol.*, **319**, 209–227.

Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graph.*, **14**, 51–55.

Kordel, J., Pearlman, D.A. and Chazin, W.J. (1997) *J. Biomol. NMR*, **10**, 231–776.

Kuboniwa, H., Tjandra, N., Grzesiek, S., Ren, H., Klee, C.B. and Bax, A. (1995) *Nat. Struct. Biol.*, **2**, 768–776.

Losonczi, J.A., Andrec, M., Fischer, M.W.F. and Prestegard, J.H. (1999) *J. Magn. Reson.*, **138**, 334–342.

McCormack, E. and Braam, J. (2003) *New Phytol.*, **159**, 585–598.

Nelson, M.R., Thulin, E., Fagan, P.A., Forsén, S. and Chazin, W.J. (2002) *Protein Sci.*, **11**, 198–205.

Ottiger, M., Delaglio, F. and Bax, A. (1998) *J. Magn. Reson.*, **131**, 373–378.

Rashidi, H.H., Bauer, M., Patterson, J. and Smith, D.W. (1999) *J. Mol. Microbiol. Biotechnol.*, **1**, 175–182.

Schwieters, C.D., Kuszewski, J.J., Tjandra, N. and Clore, G.M. (2003) *J. Magn. Reson.*, **160**, 65–73.

Skelton, N.J., Kordel, J., Akke, M., Forsén, S. and Chazin, W.J. (1994) *Nat. Struct. Biol.,* **1**, 239–245.

Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR*, **4,** 171–180.

Zhao, Q., Frederick, R.O., Seder, K.D., Thao, S., Sreenath, H., Peterson, F.C., Volkman, B.F., Markley, J.L. and Fox, B.G. (2004) *J. Struct. Funct. Genom.*, in press.

Zweckstetter, M. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 3791–3792.