Electronic Supplementary Material (ESI) for Chemical Communications.

Supporting Information

Binding and Backbone Dynamics of Protein under Topological Constrain: Calmodulin as a Model System

Priya Katyal, Yongkun Yang, You-Jun Fu, Jennifer Iandosca, Olga Vinogradova*, Yao Lin*

*To whom correspondence should be address.

Email: yao.lin@uconn.edu and olga.vinogradova@uconn.edu

1. Abbreviations:

CaM: Calmodulin

smMLCK: smooth muscle Myosin Light Chain Kinase

ITC: Isothermal Transition Calorimetry

NMR: Nuclear Magnetic Spectroscopy

NOE: Nuclear Overhauser Effect

HSQC: Heteronuclear Single Quantum Coherence

MS: Mass Spectrometry

DSC: Differential Scanning Calorimetry

β -ME: beta-mercaptoethanol

2. Methods:

Protein Expression and Purification. The gene encoding modified CaM was ordered from GenScript, in which the His_6-tag, thrombin cleavage site and an aldehyde-tag (LCTPSR) were placed at the N-terminus of calmodulin (residues 2-149, Uniprot P62161). Another aldehyde-tag (LCTPSR) was placed at the C-terminus just before the stop codon. The gene was inserted into the pET-28b vector (Novagen) between the NcoI and XhoI restriction sites. The resulting plasmid was transformed in the BL21(DE3) cells (Invitrogen). The cells were incubated in LB media with kanamycin with shaking at 37 °C until OD_{600} = 0.6. The temperature was lowered to 18 °C, and 100 μ M IPTG was added to induce expression of the aldehyde-tagged CaM. The cells were harvested by centrifugation after 12-16h. His_6-tagged CaM was purified using Ni-agarose resin (Thermo Scientific) in Tris buffer, pH 8, 300 mM NaCl. Purified CaM was then eluted using a concentration gradient of imidazole (concentration ranging from 5 mM to 1 M). Eluates were subjected to SDS-PAGE analysis and fractions containing CaM were pooled and extensively dialyzed using 20 mM Tris, pH 8. The dialyzed sample was further purified using anion exchange chromatography (Resource Q) and the peak corresponding to CaM was further purified through HiLoad 16/60 Superdex 75 column (GE Healthcare) equilibrated with 100 mM KCl, pH 6.3.

Isothermal Titration Calorimetry (ITC). Isothermal titration calorimetry was performed on a low volume Nano ITC, manufactured by TA instruments, USA. The (smMLCKp) peptide ARRKWQKTGHAVRAIGRLSS, corresponding to the calmodulin binding region of smooth muscle myosin light chain kinase, was synthesized by Applied Biosystems, Inc., and further purified by reverse phase HPLC. All experiments were carried out in 20 mM phosphate buffer, pH 6.3, 100 mM KCl, 6.1

mM $CaCl_2$ at 35°C with a stirring speed of 200 rpm. Each titration consisted of twenty 2.5 μl injections with 300 s time intervals. As a control, the heat of dilution was also measured by injecting peptide into buffer solution containing no CaM. The heat of binding was obtained as the difference between the heat of reaction and the corresponding heat of dilution. Data analysis was done in NanoAnalyze Software suite using an "independent" model. In all cases, a stoichiometry of 1 ± 0.1 was revealed.

Differential Scanning Calorimetry (DSC). The purified linear and cyclic CaM at concentrations of 0.3 mM were used for the experiments. The samples were referenced against the buffer obtained as a filtrate from the concentration step. Prior to the run, the samples were degassed using vacuum degasser. Thermograms were obtained by scanning from 10 to 130°C at 1°C/h (scanning rate) at a constant pressure of 3 atm. In total, three sets of heating, cooling cycles were done using NanoDSC (TA Instruments, USA). The data was corrected for baseline with a thermogram obtained from the buffer scans.

Nuclear Magnetic Resonance Spectroscopy. To produce ^{13}C and ^{15}N isotopically labeled CaM, cells were grown in M9 minimal media containing ^{13}C glucose (2.5g/ltr) and $^{15}NH4Cl$ (1.1g/ltr) as the sole source of carbon and nitrogen. NMR samples were prepared as described in Barbato et al.[2] Briefly, a 1mM CaM was prepared in 95% H2O/5% D_2O, 6.1 mM $CaCl_2$, and 100 mM KCl, pH 6.3. The backbone ^{1}H, ^{13}C, and ^{15}N resonances were assigned based on triple-resonance experiments: HNCA, HNCO, HNCACB, HN(CO)CA, HN(CA)CO and CBCA(CO)NH, collected using Varian/Agilent BioPack sequences. Heteronuclear–^{15}N NOEs, longitudinal (R1), and transverse (R2) ^{15}N relaxation rates were measured using standard two-dimensional methods.[3] Steady-state hetero-nuclear ^{1}H-^{15}N NOE values were determined from spectra recorded with 3s relaxation delay and in the presence and absence of a proton presaturation period of 4s. ^{15}N T1 values were measured from the spectra recorded with 7 different durations relaxation delays of T= 0.03, 0.08, 0.14, 0.3, 0.5, 0.85 and 1 ms. ^{15}N T2 values were determined from the spectra recorded with 7 different durations of the delay: T= 0.01, 0.03, 0.05, 0.09, 0.13, 0.170 and 0.25 ms. T1, T2 and NOE values were extracted by a curve-fitting subroutine included in the CCPN software suite.[4] All experiments were performed at 35°C on 600 MHz magnet (Agilent, USA) equipped with inverse-triple resonance cold probe. The generalized order parameters (S^2) and rotational correlation times (τ_c) were obtained through the FAST-ModelFree program.[6] The data was analysed using isotropic diffusion tensor and residues that had been initially fit with S^2 values less than 0.7 were excluded from the optimization of the overall rotational correlation times (τ_c). Since relaxation data was collected using single 600 MHz field, we are confident about models 1, 2 and 3, while three parameters models (4 and 5) cannot be unambiguously defined because the number of degrees-of-freedom is zero. The residues showing spectral overlap in the ^{15}N-HSQC and those associated with erroneous values in model-free

fitting (SSE cutofff 0.95 and FTest cutoff 0.8; the error >10%) were also omitted from the relaxation analysis. For residues where models 2 and 3 were chosen, the te or Rex is listed in Table S1 for l-CaM and c-CaM. All spectra were processed with NMRPipe[5] and analyzed by CCPN software suite.

LC-MS/MS analysis. The purified protein was reconstituted in 20 mM ammonium bicarbonate, and digested using trypsin in solution following either a protocol with a reducing and alkylating step or a protocol without the reducing and alkylating step.

3. Supplemental Figures (S1-S5):

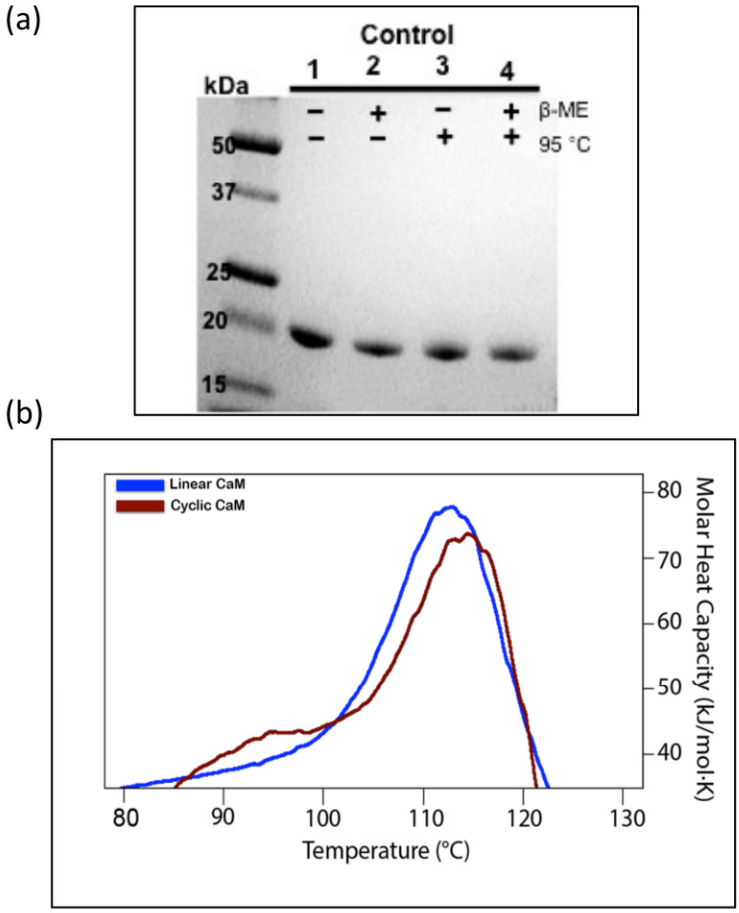

Figure S1. (a) SDS-PAGE of l-CaM under different sample conditions. (b) DSC thermograms of l-CaM (in blue line) and c-CaM (in red line)

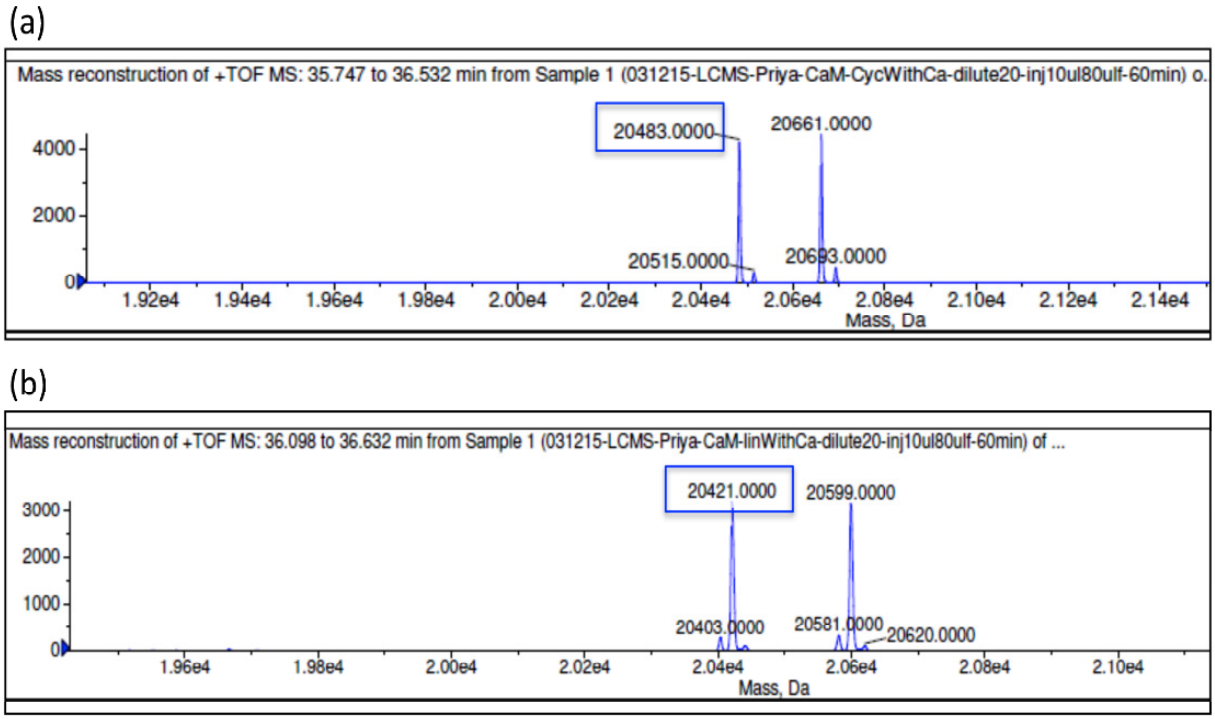

Figure S2. Reconstructed MS spectra of c-CaM (a) and l-CaM (b). The observed mass (boxed peaks) is in consistent with calculated molar mass, with c-CaM = 20.486 kDa, and l-CaM = 20.422 kDa, respectively. The second peak having a mass of M+178 m/z represents phosphogluconylation of His-Tag, commonly observed in fusion proteins (Geoghegan, K. F. *et al.* 1999)[1].

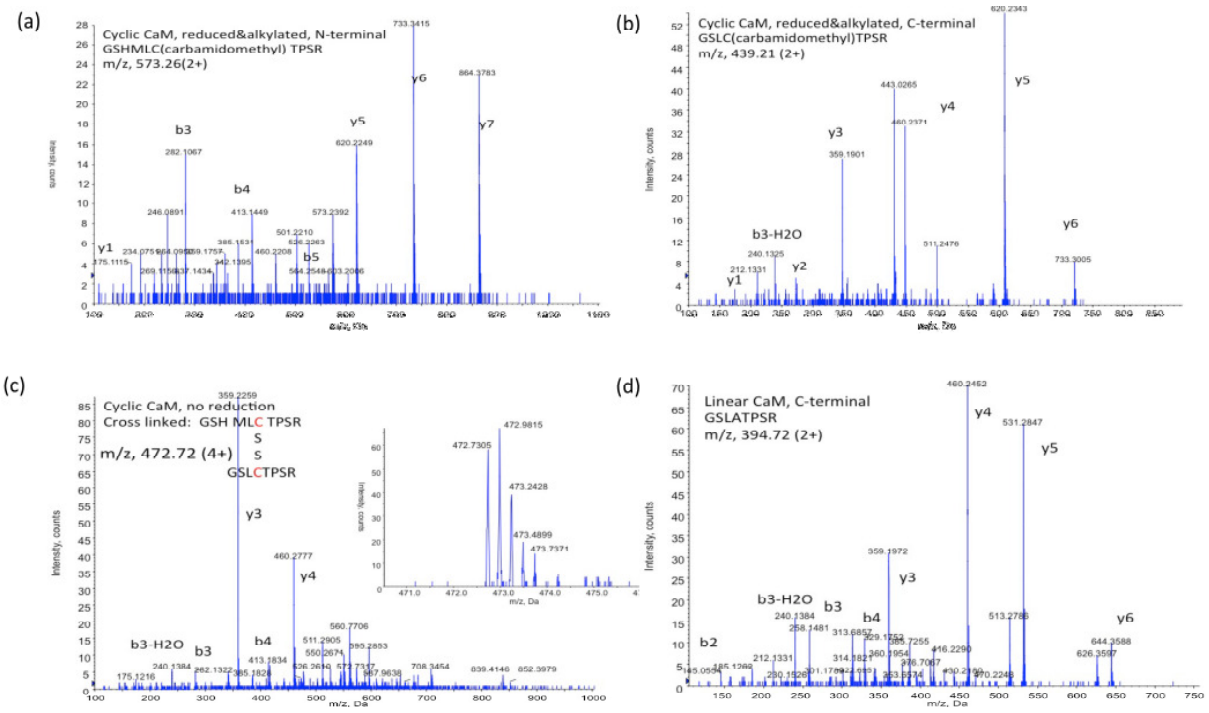

Figure S3. MS/MS spectra of c-CaM (**a-c**) and l-CaM (**d**). The samples were reduced and alkylated by DTT (dithiothreitol) and IAA (iodoacetamide) before trypsin digestion. Tryptic peptide ions, the N-terminus GSHMLC(carbamidomethyl) TPSR (m/z, 573.26 (2+)) and C-terminus GSLC(carbamidomethyl)TPSR (m/z, 439.21 (2+)), were readily identified from the relative MS/MS spectra. l-CaM bearing by LATPSR sequence generated tryptic peptide ions, N-terminus GSHMLATPSR (m/z, 528.76 (2+)) and C-terminus GSLATPSR (m/z, 394.72 (2+)).

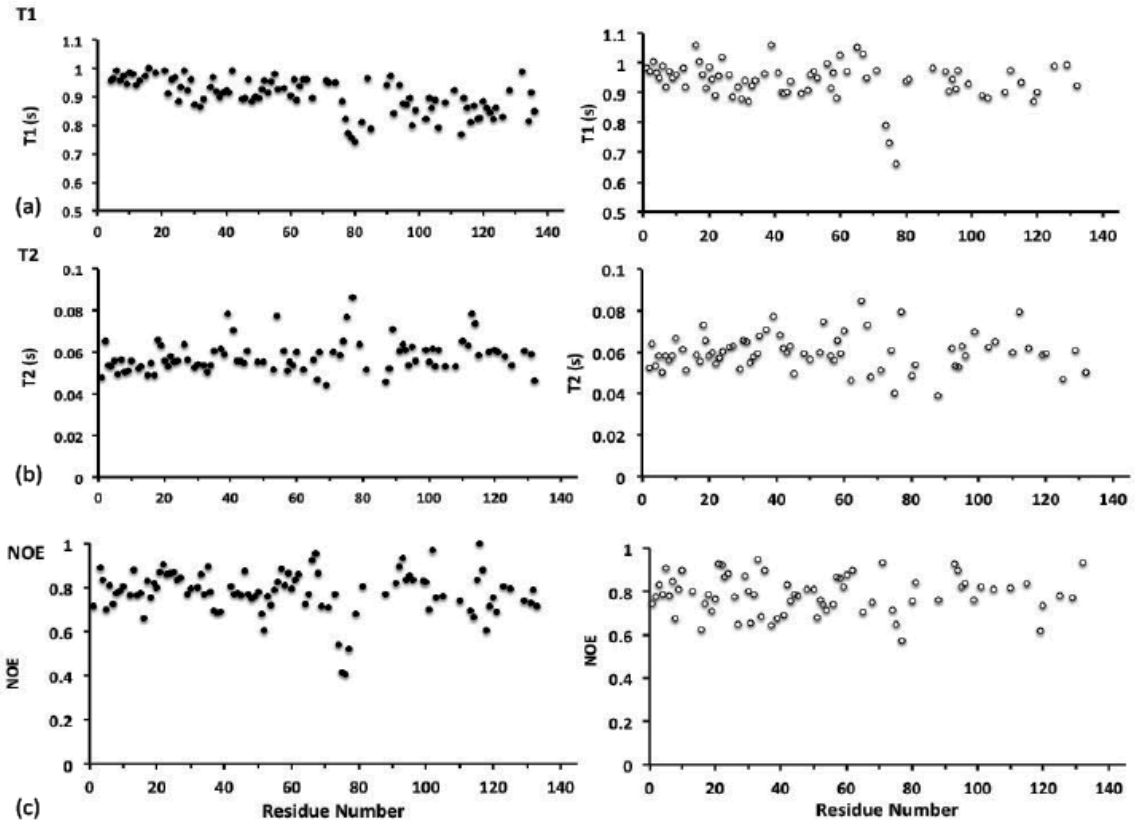

Figure S4. (a)-(c) Comparison of the ^{15}N relaxation time, T1, T2, and heteronuclear NOE measured at 600 MHz for the l-CaM (closed circles) and c-CaM (open circles).

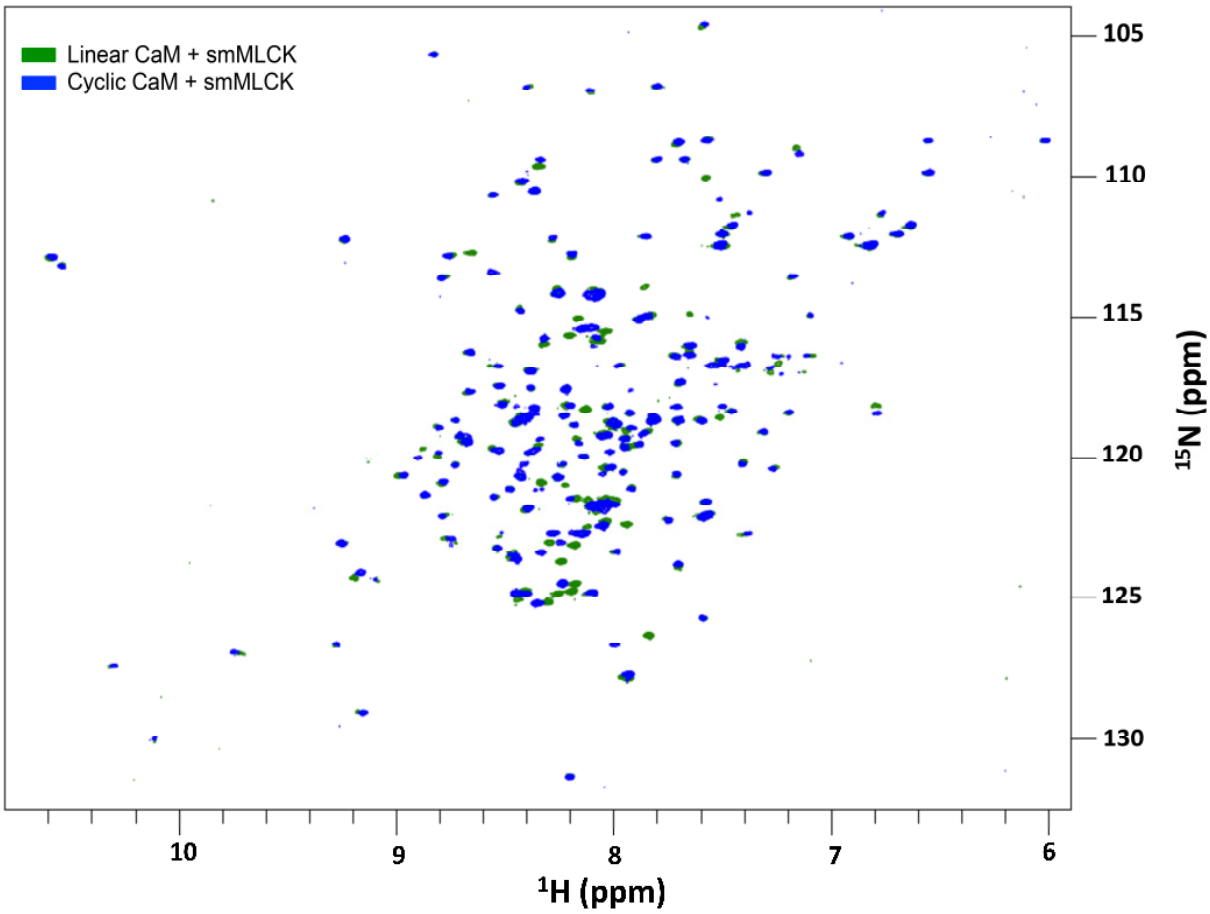

Figure S5. The superimposition of ^{15}N-HSQC spectra of l-CaM (shown in green) and c-CaM (shown in blue) in the presence of smMLCK peptide at a molar ratio of 1:2.

4. Supplemental Tables (S1-S3):

Table S1. List of residues of l-CaM exhibiting Models 2 and 3 spins

Residue	Model	te	te(err)	Rex	Rex (err)
Model 2 spins:					
41	2	1.8879e+02	2.4741e+02		
74	2	8.2711e+01	1.4493e+02		
77	2	8.9047e+02	1.4147e+02		
78	2	9.5455e+02	8.1853e+01		
82	2	1.2942e+03	2.6397e+02		
104	2	1.0328e+03	3.5911e+02		
106	2	2.0000e+03	2.7329e+02		
113	2	1.8155e+03	2.4186e+02		
121	2	7.0876e+02	2.5619e+02		
123	2	1.6763e+03	3.6031e+02		
124	2	9.6309e+02	3.5629e+02		
126	2	2.0000e+03	4.3375e+02		
134	2	1.6637e+03	3.5825e+02		
Model 3 spins:					
9	3			2.980	0.874
11	3			3.169	0.799
18	3			3.883	0.824
36	3			3.008	0.828
65	3			2.523	0.803
68	3			2.870	0.803
84	3			2.480	0.815
118	3			0.000	0.490
135	3			3.886	0.736

Table S2. List of residues of c-CaM exhibiting Models 2 and 3 spins

Residue	Model	te	te(err)	Rex	Rex (err)
Model 2 spins:					
30	2	5.7122e+02	2.8242e+02		
34	2	7.6142e+01	4.2470e+01		
44	2	4.5279e+01	1.6518e+01		
77	2	1.6402e+03	3.6893e+02		
80	2	1.4882e+03	1.4889e+02		
122	2	3.8324e+02	2.3359e+02		
Model 3 spins:					
5	3			3.067	0.773
7	3			2.656	0.767
9	3			4.283	0.813
16	3			2.634	0.744
32	3			2.407	0.805
48	3			3.633	0.811
65	3			5.546	0.886
71	3			4.363	0.858
83	3			4.099	0.856
84	3			2.119	0.762
91	3			9.852	0.907
128	3			5.580	0.804

Table S3. Thermodynamic parameters of smMLCK binding to l-CaM and c-CaM from fitting of the ITC profiles shown in Fig. 4

Thermodynamic factor	Linear CaM (kJ/mol)	Cyclic CaM (kJ/mol)
ΔG	-39.1 ± 0.1	-40.0 ± 0.7
ΔH	-55.6 ± 1.9	-18.8 ± 5.3
$-T\Delta S$	16.5 ± 1.7	-21.1 ± 4.5

5. References:

1. K. F. Geoghegan, H. B. Dixon, P. J. Rosner, L. R. Hoth, A. J. Lanzetti, K. A. Borzilleri, E. S. Marr, L. H. Pezzullo, L. B. Martin, P. K. LeMotte, A. S. McColl, A. V. Kamath and J. G. Stroh, *Anal Biochem*, 1999, **267**, 169-184.
2. G. Barbato, M. Ikura, L. E. Kay, R. W. Pastor and A. Bax, *Biochemistry*, 1992, **31**, 5269-5278.
3. L. E. Kay, D. A. Torchia and A. Bax, *Biochemistry*, 1989, **28**, 8972-8979.
4. W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue, *Proteins*, 2005, **59**, 687-696.
5. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax, *Journal of biomolecular NMR*, 1995, **6**, 277-293.
6. R. Cole and J. P. Loria, *Journal of biomolecular NMR*, 2003, **26**, 203-213.