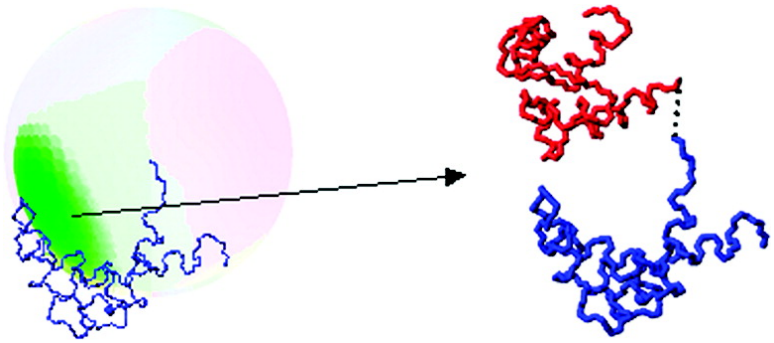
Article

Paramagnetism-Based NMR Restraints Provide Maximum Allowed Probabilities for the Different Conformations of Partially Independent Protein Domains

Ivano Bertini, Yogesh K. Gupta, Claudio Luchinat, Giacomo Parigi, Massimiliano Peana, Luca Sgheri, and Jing Yuan

J. Am. Chem. Soc., **2007**, 129 (42), 12786-12794• DOI: 10.1021/ja0726613 • Publication Date (Web): 02 October 2007



JACS
ARTICLES
Published on Web 10/02/2007

Paramagnetism-Based NMR Restraints Provide Maximum Allowed Probabilities for the Different Conformations of Partially Independent Protein Domains

Ivano Bertini,*[,†,‡] Yogesh K. Gupta,[†] Claudio Luchinat,[†,§] Giacomo Parigi,[†,§] Massimiliano Peana,[†] Luca Sgheri,[||] and Jing Yuan[†]

Contribution from the Magnetic Resonance Center (CERM), University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy, Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy, Department of Agricultural Biotechnology, University of Florence, via Maragliano 75-77, 50144 Florence, Italy, and Istituto per le Applicazioni del Calcolo—Sezione di Firenze, Polo Scientifico—CNR, Via Madonna del Piano 10, 50019 Sesto Fiorentino, Italy

Received April 17, 2007; E-mail: ivanobertini@cerm.unifi.it

Abstract: An innovative analytical/computational approach is presented to provide maximum allowed probabilities (MAPs) of conformations in protein domains not rigidly connected. The approach is applied to calmodulin and to its adduct with α-synuclein. Calmodulin is a protein constituted by two rigid domains, each of them composed by two calcium-binding EF-hand motifs, which in solution are largely free to move with respect to one another. We used the N60D mutant of calmodulin, which had been engineered to selectively bind a paramagnetic lanthanide ion to only one of its four calcium binding sites, specifically in the second EF-hand motif of the N-terminal domain. In this way, pseudocontact shifts (pcs's) and self-orientation residual dipolar couplings (rdc's) measured on the C-terminal domain provide information on its relative mobility with respect to the domain hosting the paramagnetic center. Available NMR data for terbium(III) and thulium(III) calmodulin were supplemented with additional data for dysprosium(III), analogous data were generated for the α-synuclein adduct, and the conformations with the largest MAPs were obtained for both systems. The MAP analysis for calmodulin provides further information on the variety of conformations experienced by the system. Such variety is somewhat reduced in the calmodulin−α-synuclein adduct, which however still retains high flexibility. The flexibility of the calmodulin−α-synuclein adduct is an unexpected result of this research.

Introduction

Conformational flexibility is a crucial feature in the mechanism of action of a number of proteins/enzymes.[1] Yet, detailed information on the conformational flexibility may be difficult to obtain.[2−5] There are proteins composed of domains that have a well-defined structure that are connected by a flexible linker, for which no information is available on the relative motion of the two domains. In some cases, such motions are critical to the function of the protein. In essence, we still lack the basic tools for understanding the relative position of the domains that can be experienced, the relative weight of each conformation, and the time scale of the motions involved. X-ray techniques may not be fully informative, because crystals may not form or, if a crystal is formed, only one "frozen" protein conformation is often observed. On the other hand, NMR techniques have long been used to obtain precious information on the mobility of the investigated systems.[2,3,6−22] However, standard techniques used to investigate mobility may not provide information on

† Magnetic Resonance Center (CERM), University of Florence.
‡ Department of Chemistry, , University of Florence.
§ Department of Agricultural Biotechnology, University of Florence.
|| Istituto per le Applicazioni del Calcolo—Sezione di Firenze.

(1) Huang, Y. J.; Montelione, G. T. *Nature* **2005**, *438*, 36−37.
(2) Ishima, R.; Torchia, D. A. *Nat. Struct.Biol.* **2000**, *7*, 740−743.
(3) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128−132.
(4) Mittermaier, A.; Kay, L. E. *Science* **2006**, *312*, 224−228.
(5) Fischer, M. W. F.; Zeng, L.; Majumdar, A.; Zuiderweg, E. R. P. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8016−8019.
(6) Szyperski, T.; Luginbuhl, P.; Otting, G.; Güntert, P.; Wüthrich, K. *J. Biomol. NMR* **1993**, *3*, 151−164.
(7) Barbato, G.; Ikura, M.; Kay, L. E.; Pastor, R. W.; Bax, A. *Biochemistry* **1992**, *31*, 5269−5278.
(8) Volkov, A. N.; Worrall, J. A. R.; Holtzmann, E.; Ubbink, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18945−18950.
(9) Fischer, M. W.; Losonczi, J. A.; Weaver, J. L.; Prestegard, J. H. *Biochemistry* **1999**, *38*, 9013−9022.
(10) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Bruschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098−6107.
(11) Tolman, J. R.; Al-Hashimi, H. M.; Kay, L. E.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123*, 1416−1424.
(12) Clore, G. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9021−9025.
(13) Jain, N. U.; Wyckoff, T. J. O.; Raetz, C. R. H.; Prestegard, J. H. *J. Mol. Biol.* **2004**, *343*, 1379−1389.
(14) van Dijk, A. D. J.; Boelens, R.; Bonvin, A. M. J. J. *FEBS J.* **2005**, *272*, 293−312.
(15) Baber, J. L.; Szabo, A.; Tjandra, N. *J. Am. Chem. Soc.* **2001**, *123*, 3953−3959.
(16) Mulder, F. A. A.; Mittermaier, A.; Hon, B.; Dahlquist, F. W.; Kay, L. E. *Nat. Struct. Biol.* **2001**, *8*, 932−935.
(17) Eisenmesser, E. Z.; Bosco, D. A.; Akke, M.; Kern, D. *Science* **2002**, *295*, 1520−1523.
(18) Bruschweiler, R. *Curr. Opin. Struct. Biol.* **2003**, *13*, 175−183.
(19) Ryabov, Y. E.; Fushman, D. *J. Am. Chem. Soc.* **2007**, *129*, 3315−3327.
(20) Ryabov, Y. E.; Fushman, D. *Magn. Reson. Chem.* **2006**, *44*, S143−S151.

10.1021/ja0726613 CCC: $37.00

the conformational space sampled by the protein. Of course, it is easy to understand that a complete description of such motions will never be obtained, because the number of experimental data is far smaller than the number of unknowns to be determined.

Similar problems in related fields such as liquid crystals have been tackled in the past using maximum entropy methods.[23−25] However, the information obtained is scarce even for systems with low complexity, unless an "a priori" physical model is imposed on the system. Investigations with the same objective of describing the preferred conformations experienced by the protein have been also performed on unfolded proteins using paramagnetic relaxation enhancements induced by spin labels[26] or residual dipolar couplings arising in the presence of orienting media in solution.[27]

Paramagnetic metal ions may provide additional NMR parameters such as pseudocontact shifts (pcs's),[28] in addition to residual dipolar couplings (rdc's) due to self-orientation of the paramagnetic molecule in high magnetic fields.[28] Such parameters may help in elucidating the long-range spatial relationships and the dynamics in proteins[29,30] and in protein−protein interactions.[31−35] Recently, NMR measurements on paramagnetic systems allowed us to obtain information on the preferred region of space experienced by one domain with respect to the other in the two-domain protein calmodulin.[36] The information contained in the pcs's and/or the rdc's was shown to be useful, as the measured values are given by the average of the values corresponding to the experienced conformations, and the two observables average very differently. Both pcs's and rdc's average when the motions occur on time scales faster than, or of the order of, 10^{-2} s and are thus able to incorporate information on motions within a very broad time scale. Pcs and rdc restraints are obtainable for several paramagnetic metalloproteins, for metalloproteins where a native diamagnetic metal ion is substituted by an appropriate paramagnetic one,[36−39] or for proteins where a paramagnetic tag is artificially attached.[35,40−43] It should be noted that rdc's induced by external devices are not useful, as the protein domains will be largely oriented by their individual interactions with the external device, while the contribution from the orientation of a nearby domain may be small in the presence of sizable motional freedom.

An innovative approach for determining the maximum allowed probability (MAP) of any conformation in a protein constituted by domains not rigidly connected is presented here. It is based on the characterization of the conformational space in terms of a maximum probability value, as defined in a recent theoretical work,[44] that is allowed, for any conformation, to be consistent with the experimental average pcs and rdc data. This MAP value is not the probability of finding the protein in that conformation but rather tells us that such a conformation cannot have a probability larger than that value. Even so, the result is quite informative. The approach is applied to a variant of calmodulin (CaM, N60D mutant) as well as to its adduct with α-synuclein (CaM-AS). CaM is a protein constituted by two rigid domains (called N-terminal and C-terminal domains) whose relative orientation is not fixed. Each domain, composed by two EF-hand motifs connected with a loop, contains two calcium binding sites, so that CaM binds up to four calcium ions in total. The N60D protein mutant had been engineered to selectively bind a paramagnetic lanthanide ion to only one of its four calcium binding sites, specifically in the second EF-hand motif of the N-terminal domain (see Figure 1).[39] Pcs and rdc data relative to two lanthanide ions (Tb^{3+} and Tm^{3+}) are already available in the literature.[36] The CaM−AS adduct was also investigated. AS is a small cytoplasmic protein (15 kDa) that is essentially unfolded in its soluble, monomeric state[45,46] and is abundant in the presynaptic space. It had been shown that monomeric AS interacts with CaM, with reported dissociation constants of the order of 10−100 nM.[47,48] The NMR data obtained here indicate that an adduct is actually formed, but with a dissociation constant in the micromolar range, therefore questioning its physiopathological relevance. On the other hand, it is found that the adduct is highly flexible, involving fast rearrangement of the relative position of the two CaM domains. This makes the CaM−AS adduct an ideal test case for our approach.

The approach is based on the measurements of pcs's of the N-terminal domain of CaM for three lanthanide derivatives, i.e., Tb^{3+}, Tm^{3+}, and Dy^{3+}, in order to determine the magnetic

(21) Ferrage, F.; Pelupessy, P.; Cowburn, D.; Bodenhausen, G. *J. Am. Chem. Soc.* **2006**, *128*, 11072−11078.
(22) Pfeiffer, S.; Fushman, D.; Cowburn, D. *J. Am. Chem. Soc.* **2001**, *123*, 3021−3026.
(23) Catalano, D.; Di Bari, L.; Veracini, C. A.; Shilstone, G. N.; Zannoni, C. *J. Chem. Phys.* **1991**, *94*, 3928−3935.
(24) Berardi, R.; Spinozzi, F.; Zannoni, C. *J. Chem. Soc., Faraday Trans.* **1992**, *88*, 1863−1873.
(25) Berardi, R.; Spinozzi, F.; Zannoni, C. *J. Chem. Phys.* **1998**, *109*, 3742−3759.
(26) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 476−477.
(27) Bernado, P.; Bertoncini, C. W.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2005**, *127*, 17968−17969.
(28) Bertini, I.; Luchinat, C.; Parigi, G. *Progr. NMR Spectrosc.* **2002**, *40*, 249−273.
(29) Gaponenko, V.; Sarma, S. P.; Altieri, A. S.; Horita, D. A.; Li, J.; Byrd, R. A. *J. Biomol. NMR* **2004**, *28*, 205−212.
(30) Bertoncini, C. W.; Jung, Y.-S.; Fernández, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1430−1435.
(31) Guiles, R. D.; Sarma, S.; DiGate, R. J.; Banville, D.; Basus, V. J.; Kuntz, I. D.; Waskell, L. *Nat. Struct. Biol.* **1996**, *3*, 333−339.
(32) Worrall, J. A. R.; Kolczak, U.; Canters, G. W.; Ubbink, M. *Biochemistry* **2001**, *40*, 7069−7076.
(33) Gochin, M. *Structure Fold Des.* **2000**, *8*, 441−452.
(34) Diaz-Moreno, I.; Diaz-Quintana, A.; De la Rosa, M. A.; Ubbink, M. *J. Biol. Chem.* **2005** *280*, 18908−18915.
(35) Pintacuda, G.; Park, A. Y.; Keniry, M. A.; Dixonj, N. E.; Otting, G. *J. Am. Chem. Soc.* **2006**, *128*, 3696−3702.
(36) Bertini, I.; Del Bianco, C.; Gelis, I.; Katsaros, N.; Luchinat, C.; Parigi, G.; Peana, M.; Provenzani, A.; Zoroddu, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6841−6846.
(37) Bertini, I.; Luchinat, C.; Parigi, G. *Concepts Magn. Reson.* **2002**, *14*, 259−286.
(38) Bertini, I.; Donaire, A.; Jiménez, B.; Luchinat, C.; Parigi, G.; Piccioli, M.; Poggi, L. *J. Biomol. NMR* **2001**, *21*, 85−98.
(39) Bertini, I.; Gelis, I.; Katsaros, N.; Luchinat, C.; Provenzani, A. *Biochemistry* **2003**, *42*, 8011−8021.
(40) Wöhnert, J.; Franz, K. J.; Nitz, M.; Imperiali, B.; Schwalbe, H. *J. Am. Chem. Soc.* **2003**, *125*, 13338−13339.
(41) Su, X. C.; Huber, T.; Dixon, N. E.; Otting, G. *ChemBioChem* **2006**, *7*, 1599−1604.
(42) Prudencio, M.; Rohovec, J.; Peters, J. A.; Tocheva, E.; Boulanger, M. J.; Murphy, M. E.; Hupkes, H. J.; Koster, W.; Impagliazzo, A.; Ubbink, M. *Chem.−Eur. J.* **2004**, *5*, 3252−3260.
(43) Ikegami, T.; Verdier, L.; Sakhaii, P.; Grimme, S.; Pescatore, P.; Saxena, K.; Fiebig, K. M.; Griesinger, C. *J. Biomol. NMR* **2004**, *29*, 339−349.
(44) Longinetti, M.; Luchinat, C.; Parigi, G.; Sgheri, L. *Inv. Probl.* **2006**, *22*, 1485−1502.
(45) Weinreb, P. H.; Zhen, W. G.; Poon, A. W.; Conway, K. A.; Lansbury, P. T., Jr. *Biochemistry* **1996**, *35*, 13709−13715.
(46) Eliezer, D.; Kutluay, E.; Bussell, R., Jr.; Browne, G. *J. Mol. Biol.* **2001**, *307*, 1061−1073.
(47) Lee, D.; Lee, S.-Y.; Lee, E.-N.; Chang, C.-S.; Paik, S. R. *J. Neurochem.* **2002**, *82*, 1007−1017.
(48) Martinez, J.; Moeller, I.; Erdjument-Bromage, H.; Tempst, P.; Lauring, B. *J. Biol. Chem.* **2003**, *278*, 17379−17387.

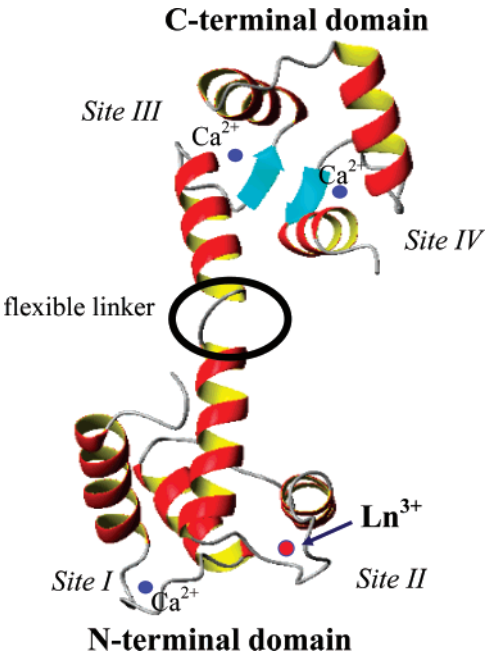


Figure 1. CaM can bind four calcium ions, two in the N-terminal domain and two in the C-terminal domain. The N60D mutant binds lanthanides selectively at the second binding site of the N-terminal domain. The two domains are shown as observed in the so-called "extended" conformation of CaM (PDB 1CLL).

susceptibility anisotropy tensors and then relate them to the conformationally averaged pcs and rdc values measured on the C-terminal domain with the different lanthanide ions. Substitution of the calcium ion with a lanthanide ion does not appreciably affect the structure of calmodulin, as shown by using the diamagnetic Lu^{3+} ion.[36,39,49] Several other EF-hand proteins are similarly well behaved.[38,50,51] The results show that for the first time it is possible to characterize the conformational space in terms of the different MAPs for each relative conformation of the two domains.

Materials and Methods

Protein Preparation. ^{15}N and ^{13}C labeled wild type and N60D CaM were purchased from ProtEra s.r.l., being expressed and purified as previously reported.[36,39] NMR samples of Ca_4CaM and $(CaLn)_N(Ca_2)_C$CaM (Ln = Tb, Tm, Dy, Lu) were prepared as previously reported.[39] Details on the preparation and purification of AS[47,48] are reported in the Supporting Information.

NMR Measurements. Labeled wild type CaM and N60D CaM were slowly titrated with unlabeled human AS. The titration progress was followed by $^1H-^{15}N$ HSQC spectra at 700 MHz and 298 K. Titrations of labeled human AS with unlabeled human CaM were performed under the same conditions.

The NMR spectra were acquired on Bruker AVANCE 600 and 700 spectrometers equipped with a triple resonance (TXI) 5 mm probe with a *z*-axis pulse field gradient. All spectra were taken at 298 K. The water signal was suppressed using presaturation during the relaxation delay and mixing time or by using the WATERGATE[52] method.

In order to obtain the pseudocontact shifts (pcs's), 298 K $^1H-^{15}N$ HSQC spectra of $(CaLn)_N(Ca_2)_C$CaM−AS were recorded. 256 increments each with 1024 complex data points and 48 transients were collected. Pcs's were calculated as the difference between the chemical shifts of corresponding nuclei in the paramagnetic and diamagnetic derivative. One bond $^1H-^{15}N$ coupling constants (rdc's) were measured at 298 K and 700 MHz by using the IPAP method.[53] In all experiments, the concentration of labeled CaM was about 0.6 mM with a slight excess of unlabeled AS; in labeled AS samples, a slight excess of unlabeled CaM was used.

Results

CaM−AS Adduct. From $^1H-^{15}N$ HSQC spectra, 146 out of 148 HN signals were observed and assigned through comparison with the spectra of the free CaM, with the help of titration with increasing amounts of AS.

The analysis of the 3D ^{13}C-edited and ^{15}N-edited NOESY-HSQC spectra of CaM in the CaM−AS sample provided the full assignment through comparison of the NOE patterns with free CaM, and 4530 intradomain NOE cross-peaks were assigned and transformed into 3288 unique upper distance limits, of which 2971 (1686 for the N-terminal domain (21.3 NOE/residue) and 1285 for the C-terminal domain (18.6 NOE/residue)) were found to be meaningful. A lower number of NOEs in the C-terminal domain has been already noted[36,54] and ascribed to some conformational averaging within that domain.[54] The structure calculations, performed with the program DYANA yielded well resolved structure families for both CaM domains.

The binding of CaM to AS was tested by following the changes in the $^1H-^{15}N$ HSQC spectrum of ^{15}N-labeled CaM upon addition of an increasing amount of unlabeled AS, up to final ratios of 1:1 (CaM−AS). Further additions of AS did not cause further appreciable changes. The chemical shifts of several peaks of CaM are affected, though slightly (Figure 2). From the titration, a dissociation constant around 10^{-5} M is estimated. Neither interdomain nor intermolecular NOEs were observed.

For AS in the CaM−AS adduct, sequential backbone connectivities were obtained as in the case for the free AS.[46,55,56] Very little shifts of either backbone or side chain signals of AS in the presence of CaM were observed.

Paramagnetism-Based Restraints in CaM and CaM−AS. Pcs and rdc data were measured for $(CaDy)_N(Ca_2)_C$CaM (see Figure 3). Pcs's and rdc's were already available for $(CaTb)_N(Ca_2)_C$CaM and $(CaTm)_N(Ca_2)_C$CaM.[36]

The same parameters were measured for $(CaLn)_N(Ca_2)_C$CaM (Ln = Tb, Tm, or Dy) in the presence of AS (see Figures 3 and 4). AS exchanges rapidly between bound and free forms and experiences very small pcs's with respect to both domains of CaM. This is presumably because AS binds CaM with different orientations. Still, it affects the conformational variability of CaM, as rdc and pcs measured for CaM in the presence of AS are different from those in free CaM.

Pcs's relative to the N-terminal domain of CaM were used to obtain the magnetic susceptibility anisotropy tensors of the three lanthanides, in addition to refining the domain structure through the program PARAMAGNETIC DYANA.[37] The tensor

(49) Bentrop, D.; Bertini, I.; Cremonini, M. A.; Forsén, S.; Luchinat, C.; Malmendal, A. *Biochemistry* **1997**, *36*, 11605−11618.
(50) Baig, I.; Bertini, I.; Del Bianco, C.; Gupta, Y. K.; Lee, Y.-M.; Luchinat, C.; Quattrone, A. *Biochemistry* **2004**, *43*, 5562−5573.
(51) Babini, E.; Bertini, I.; Capozzi, F.; Del Bianco, C.; Holleder, D.; Kiss, T.; Luchinat, C.; Quattrone, A. *Biochemistry* **2004**, *43*, 16076−16085.
(52) Piotto, M.; Saudek, V.; Sklenar, V. *J. Biomol. NMR* **1992**, *2*, 661−666.
(53) Ottiger, M.; Delaglio, F.; Bax, A. *J. Magn. Reson.* **1998**, *131*, 373−378.
(54) Kuboniwa, H.; Tjandra, N.; Grzesiek, S.; Ren, H.; Klee, C. B.; Bax, A. *Nat. Struct. Biol.* **1995**, *2*, 768−776.
(55) Fernandez, C. O.; Hoyer, W.; Zweckstetter, M.; Jares-Erijman, E. A.; Subramaniam, V.; Griesinger, C.; Jovin, T. M. *EMBO J.* **2004**, *23*, 2039−2046.
(56) Bermel, W.; Bertini, I.; Felli, I. C.; Lee, Y.-M.; Luchinat, C.; Pierattelli, R. *J. Am. Chem. Soc.* **2006**, *128*, 3918−3919.

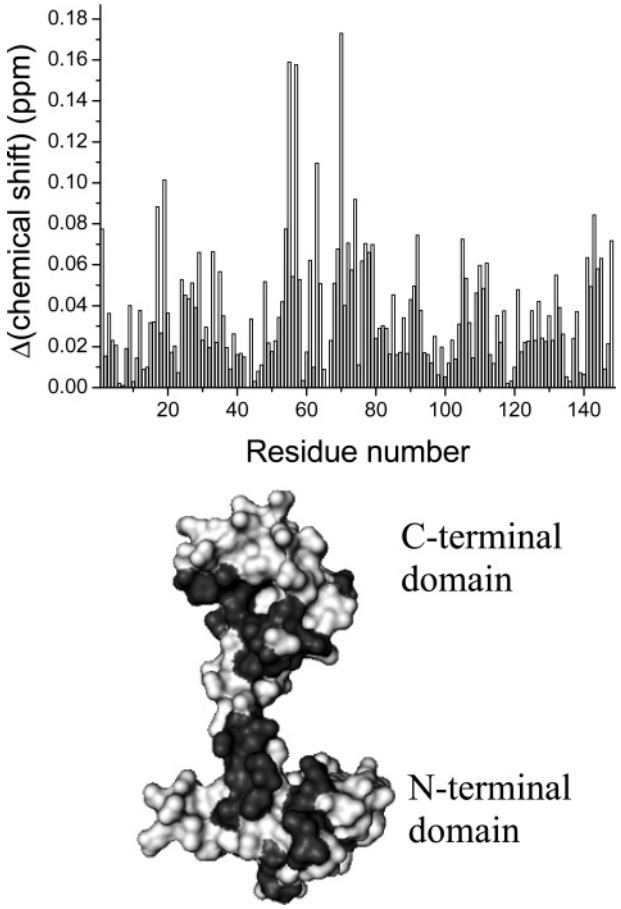


Figure 2. Plot of the change in chemical shift observed for the CaM H^N and N backbone atoms in the 1:1 ^{15}N AS−CaM complex. Shifts are reported as a weighted average of the amide proton and amide nitrogen shifts using the formula $\Delta\delta = [(\Delta\delta_H)^2 + (\Delta\delta_N \times (\{\gamma_N\}/\{\gamma_H\}))^2]^{0.5}$. Residues with $\Delta\delta$ values larger than 0.05 ppm are shown in gray on the CaM structure in the "extended" conformation observed in the solid state (PDB 1CLL).

parameters are reported in Table 1. The structure of the C-terminal domain was refined using the rdc's relative to the C-terminal domain, to make them as consistent as possible with the structure. The backbone rmsd between residues 5−72 of the family of the 20 structures with the lowest target function is 0.46 Å, and that between residues 82−143 is 0.50 Å. Both structures remain very similar to those previously reported.[36,57]

Rdc's do not depend on distance, and therefore the spreading of their values should be approximately the same in both the N- and C-terminal domains, if there were no relative motion between the two.[58] The spreading of the rdc measured in the C-terminal domain in CaM−AS is much smaller than predicted for a rigid molecule (see Figure 3) but sizably larger than that observed in the free CaM protein. It can be concluded that indeed the CaM domains in the adduct with AS are highly flexible but appreciably less so than in the free form. The small rdc values measured for CaM−AS, in fact, cannot result from the sum of the contributions from a free CaM form in chemical equilibrium with a CaM−AS form assuming a closed conformation, because the dissociation constant for the complex ensures that the CaM−AS form is surely present with a percentage larger than 90% under the present experimental conditions.

Figure 4 shows the pcs values observed for the C-terminal HN nuclei in the free and AS-bound CaM forms. The somewhat larger values measured in the AS-bound form suggest a slightly shorter average distance of the C-terminal domain from the paramagnetic metal ion located in the N-terminal domain or a smaller dynamic (orientational) averaging, possibly due to an increase in the localization of the C-terminal domain in a region of space with pcs values of the same sign as that of the experimental ones.

It appears that a single structure of the whole CaM molecule cannot be calculated, even in the presence of interaction with AS, due to its high flexibility, and therefore, this system can be used, together with the free CaM, as a test case to apply our strategy for the estimate of MAP conformations of the protein.

MAP Values. A novel approach is developed here to extract from rdc and pcs data the conformations that have the largest MAP value among all possible conformations. The maximum allowed probability of a given orientation of one domain with respect to another domain of the same protein using only rdc data was earlier defined and called p_{max}.[59] This quantity represents the maximum weight that a given *orientation* can have and does not depend on the number and weight of all the other orientations that the domain may experience. Rdc data, in fact, provide information only on orientation (determined by a rotation matrix R). In the present framework, we term this orientational MAP as MAP(R). To define a *conformational* MAP, we take MAP(R) as the starting point, to which *translational* information must be added.

The nature of rdc's (which are independent of reflections of the axes of the magnetic tensor) is such that the same MAP(R) is calculated for a given orientation as well as for other 3 symmetric orientations, or *ghost* orientations, which cannot be discriminated. In principle, two (or more) different metal ions with significantly different magnetic susceptibility anisotropy tensors and good quality rdc values should be able to eliminate the ghost orientations.[44] However, simulations performed using three paramagnetic ions which induce magnetic susceptibility tensors with similar orientation, as expected for lanthanide ions in the same binding pocket, show that the ghosts are not completely removed (see Supporting Information). Furthermore, even if further paramagnetic ions are considered, little additional information is added, because further lanthanides do not have a significantly different orientation of the magnetic anisotropy susceptibility tensor.

Since the relative position of the two domains is restricted by the presence of a physical linker, coupling rdc's with pcs's could in principle remove some of the ambiguities, in addition to providing further information on the relative position(s) of the domains. Therefore, we introduce pcs's in the analysis and we define an MAP relative to each conformation, defined by orientation plus translation. In practice, two sets of rdc's and pcs's may not completely remove all the ambiguities, because we notice that by adding a third set of data the ghost solutions

(57) Chou, J. J.; Li, S.; Klee, C. B.; Bax, A. *Nat. Struct. Biol.* **2001**, *8*, 990−997.
(58) Clore, G. M.; Gronenborn, A. M.; Bax, A. *J. Magn. Reson.* **1998**, *133*, 216−221.
(59) Gardner, R. J.; Longinetti, M.; Sgheri, L. *Inv. Probl.* **2005**, *21*, 879−898.
(60) Bertini, I.; Janik, M. B. L.; Lee, Y.-M.; Luchinat, C.; Rosato, A. *J. Am. Chem. Soc.* **2001**, *123*, 4181−4188.
(61) Barbieri, R.; Bertini, I.; Cavallaro, G.; Lee, Y.-M.; Luchinat, C.; Rosato, A. *J. Am. Chem. Soc.* **2002**, *124*, 5581−5587.

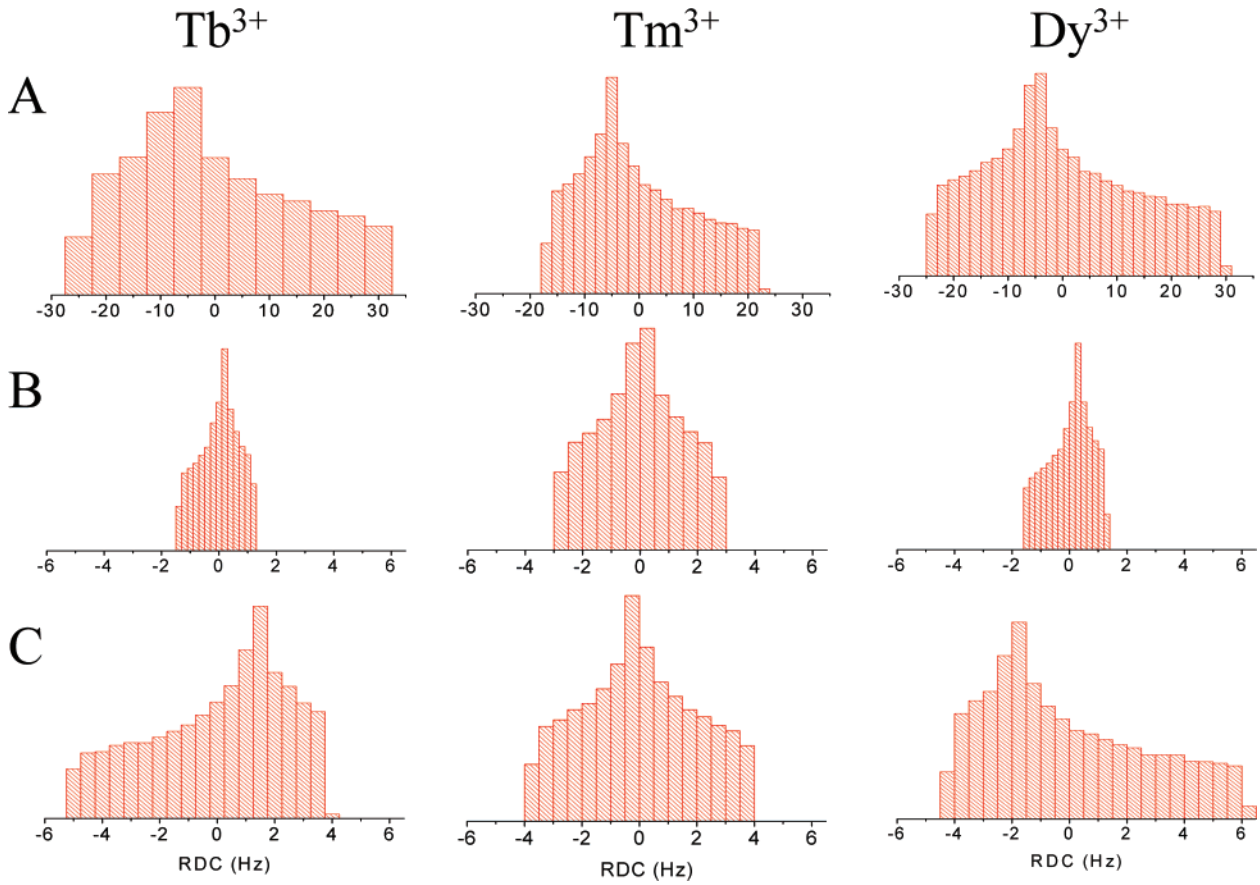


Figure 3. Observed spreading of rdc values in the C terminal domain of $(CaLn)_N(Ca_2)_C$CaM (Ln = Tb, Tm, or Dy) in the free form (B) or complexed with AS (C) as compared with the spreading predicted in the absence of conformational freedom (A).

keep disappearing. Simulations with exact data, on the other hand, show that three lanthanides remove all ghost solutions (see Supporting Information). Of course, the efficiency of pcs's in removing ghost solutions increases with the magnitude of the pcs values, as may happen in domains closer than those in CaM or in the presence of more limited conformational freedom.

Because of the different mathematical structure of the pcs and rdc equations, however, the geometric algorithm proposed in Longinetti et al.[44] (see Supporting Information) for the determination of MAP(R) presents many theoretical and practical difficulties. Therefore, we used the following procedure. The MAP(R) values relative to all orientations of one domain with respect to the other domain were first obtained using the rdc values, through the approach proposed in ref 44. Then, a fit was performed starting from selected orientations R_0 with the largest MAP(R_0) values, complemented by another N conformations, with weight (w_i), position (t_i), and orientation (R_i) obtained in order to minimize the target function

$$TF(w_0) = \min_{t_0,(w_i,t_i,R_i)} \sum_j |\tilde{\delta}_j - (w_0\delta_j(t_0,R_0) + \sum_{i=1}^{N} w_i\delta_j(t_i,R_i))|^2 \quad (1)$$

where $\tilde{\delta}_j$ are the experimental pcs/rdc values, $\delta_j(t_0,R_0)$ are the pcs/rdc values calculated for the selected orientation R_0, with the translation vector t_0 defining its position, w_0 is the corresponding weight, and $\delta_j(t_i,R_i)$ are the pcs/rdc values calculated for the other $i = 1...N$ conformations. Such a function (with $w_0 + \sum w_i = 1$) represents the minimal error on the reconstructed data when the domain is constrained to stay in orientation R_0 for a fraction w_0 of the time. A weighting factor is introduced to normalize the contributions to the target function from pcs's and from rdc's according to their squared values and to make them of the same order. The function $TF(w_0)$ is calculated for increasing values of w_0 and increases with w_0. The absolute minimum of $TF(w_0)$ is $TF(0)$, which does not depend on t_0 and R_0. Then the MAP value is the largest w_0 value such that $TF(w_0) = \epsilon$, where ϵ is the threshold fixed for the error. This was set to a 10% larger value of the absolute minimum of the TF.

A simulated annealing minimization procedure was applied for the determination of the other N conformations. Such minimization, which includes $N \times 7 - 1$ variables (3 translations, 3 rotations, and 1 weighting factor for any conformation except the last one), needs to be handled carefully. The fit protocol is reported in detail in the Supporting Information. About 2−3 days of CPU time on a single Pentium-4 3.2 GHz processor are required to provide the TF for each conformation and a fixed w_0 value. Calculations were then repeated for several w_0 weights in order to obtain the MAP value. Faster minimization procedures could be attempted, but care should be taken to extensively search all the conformational space to exclude the possibility that another set of $N + 1$ conformations would have provided a lower TF.

The achievement of an accurate estimate of the MAP requires that a high enough number $N + 1$ of conformations is considered, although the experimental averaged pcs and rdc data may be reconstructed in some cases with less conformations.

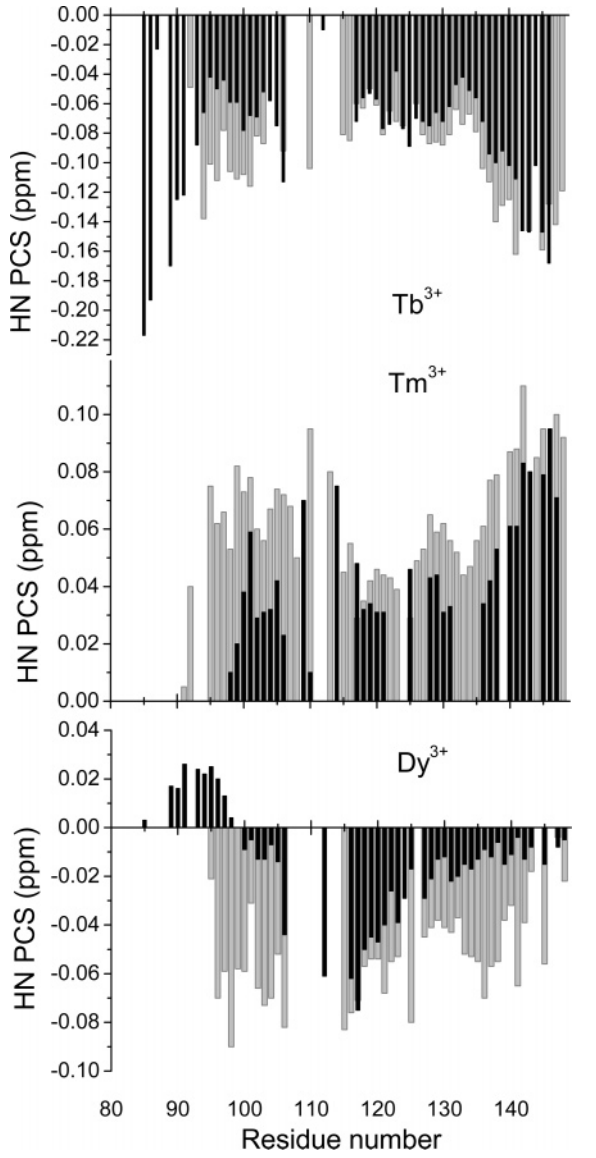


Figure 4. Observed C-terminal HN pcs values for the terbium(III), thulium(III), and dysprosium(III) CaM derivatives in the free form (black) and in the presence of AS (gray).

Table 1. Magnetic Susceptibility Anisotropies of the Different Lanthanides in CaM and CaM−AS

	$\Delta\chi_{ax}$ (10^{-32} m^3)	$\Delta\chi_{rh}$ (10^{-32} m^3)	Euler angles[a] (referring to PDB 1J7O, rad)
$(CaTb)_N(Ca_2)_CCaM$	37	−14	1.828 1.246 0.248
$(CaDy)_N(Ca_2)_CCaM$	34	−15	1.208 0.323 0.672
$(CaTm)_N(Ca_2)_CCaM$	26	−9.1	0.232 −1.953 −0.324
$(CaTb)_N(Ca_2)_CCaM-AS$	33	−17	1.665 1.053 0.571
$(CaDy)_N(Ca_2)_CCaM-AS$	31	−13	1.204 0.282 0.654
$(CaTm)_N(Ca_2)_CCaM-AS$	23	−9.3	0.200 −1.890 −0.250

[a] Defined as yaw, roll, and pitch. The magnetic susceptibility anisotropy values are similar to those observed in other EF-hand proteins.[50,51,60] The spread in the directions of the principal axes of the χ tensors of the three metals is large enough to consider the three datasets independent from one another, as previously observed.[36,61]

The number $N+1$ of conformations to be used to achieve the absolute minimum for the $TF(w_0)$ (which is independent of N for large enough values) depends on the number of different metal ions, m, to which pcs and rdc data refer. Actually, this number is theoretically limited to $5m$ as far as rdc's are concerned, whereas it can reach the number of available

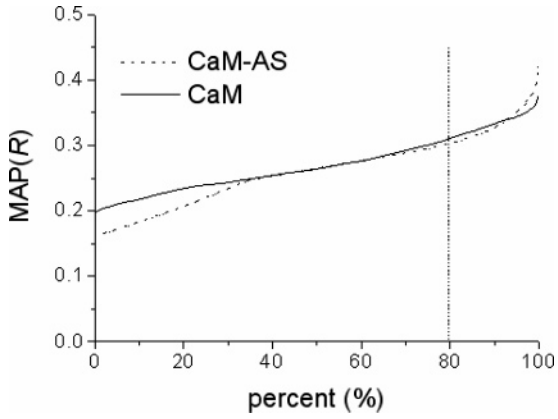


Figure 5. MAP(R) values calculated from rdc data for CaM and CaM−AS. A point (x,y) on the graph means that a fraction x of all orientations have a value of MAP(R) $\leq y$.

restraints, as far as pcs's are concerned. In practice a much smaller value N is usually needed. For the present calculations, where restraints relative to three metal ions were employed, N was fixed to 9. We verified that the addition of further conformations did not decrease the target function and, thus, could not increase the MAP of the fixed conformation. Furthermore, no analytical cases were found requiring more than 8 conformations to reproduce rdc data corresponding to three metals; pcs's, on the other hand, can be easily fit in our case. In fact they provide, when taken alone, quite large MAP values for all conformations, so that the remaining weight (1 − MAP) is small and the number of conformations needed to accommodate it is small as well. Furthermore, when rdc's and pcs's are taken together, the pcs's can be mostly accommodated using the translations, which do not change the rdc's.

Finally, synthetic tests were performed by modeling the location of the C-terminal domain with respect to the N-terminal domain in a wide range of orientations. A very large number (50 000) of protein conformations were generated using a Gaussian probability distribution around one selected conformation. Rdc and pcs data were simulated from the average of rdc's and pcs's obtained for the different conformations. They were then used according to the proposed procedure. Calculations performed using pcs's and rdc's relative to 3 or 5 metal ions indicate that the conformations with the largest MAP are close to the center of the Gaussian distribution used to generate the data. Such agreement is maintained when a stochastic error is introduced (±30% for pcs's, ±0.5 Hz for rdc's). Details on the tests performed are reported in the Supporting Information.

Determination of the Largest MAP Values for CaM and CaM-AS. The algorithm described above was applied to monitor the conformational space sampled by CaM and CaM−AS, using the pcs and rdc data measured for Tb^{3+}, Tm^{3+}, and Dy^{3+}. Figure 5 shows the MAP(R) values calculated from rdc data only, and Figures 6 and 7 show the conformations with the largest (≥0.35) MAP(R) and MAP values. In Figures 6 and 7 the points on the sphere and their colors (see below) represent the preferential relative orientations of the C-terminal domain with respect to the N-terminal domain when the first residue of the former and the last residue of the latter are both placed in the center of the sphere. In this way, it is possible not only to visualize the most probable orientations of the C-terminal

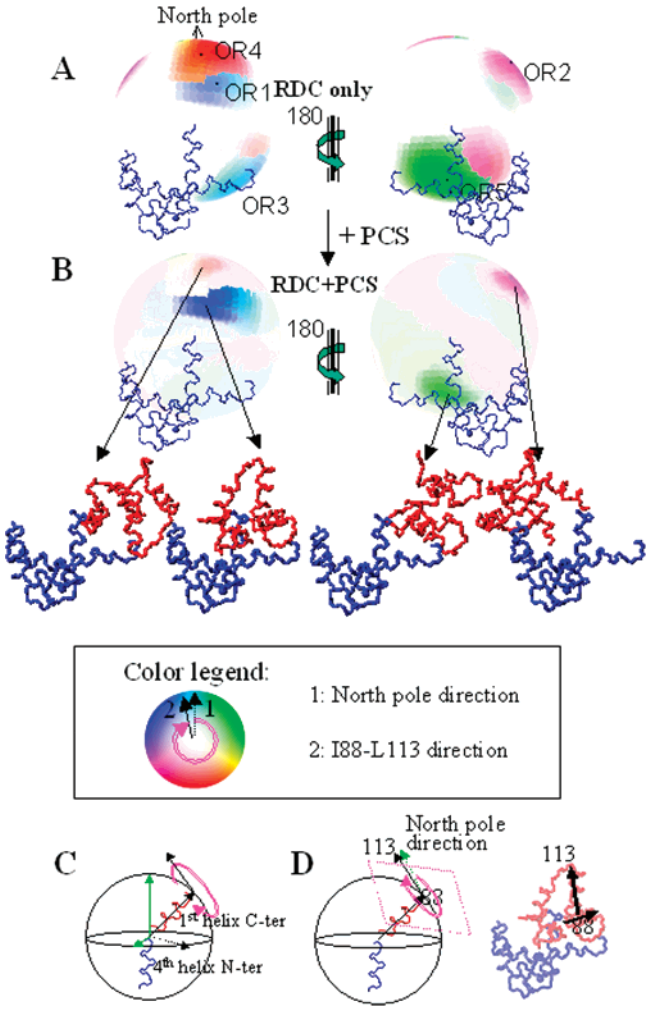


Figure 6. (A) MAP(*R*) values calculated from rdc data for all relative orientations of the C-terminal domain with respect to the N-terminal domain (in blue) of free CaM. The first C-terminal residue and the last N-terminal residue outside the mobile 78−81 hinge region are placed in the center of the sphere. The points on the sphere represent the directions of the first helix on the C-terminal domain (C). The colors represent the angle between the projections of the vector connecting the atoms C′ of residue Ala 88 and C^{α} of residue Gly 113 (virtually perpendicular to the axis of the first helix of the C-terminal domain) and the projection of the North pole direction, on the plane tangent to the sphere in each point, according to the legend (D). (B) Conformations with the largest MAP values in agreement with both rdc and pcs data. The intensity of the color is low for conformations with MAP(*R*) or MAP < 0.35 and increases proportionally with increasing MAP(*R*) or MAP above that threshold.

domain but also to figure out the most probable *conformations* of the whole protein with the assumption that the translational displacement between the end of the last helix of the N-terminal domain and the beginning of the first helix of the C-terminal domain is modest.

As shown in Figure 6 for free CaM, the orientations of the C-terminal domain with respect to the N-terminal domain are defined by three angles: two of them provide the orientation of the first helix of the C-terminal domain, the third one describes the rotation of the C-terminal domain around its first helix (see Figure 6C). The first two angles thus define points on a sphere in correspondence of the direction of the first helix of the C-terminal domain. The value of the third angle is depicted according to a color code (see the color legend in Figure 6). The colors represent the angle between a vector chosen in the plane perpendicular to the first helix of

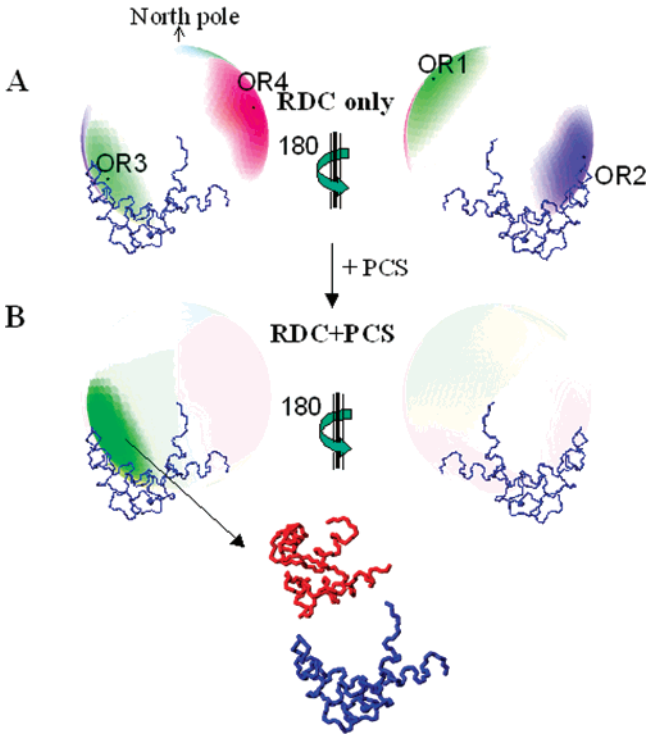


Figure 7. (A) MAP(*R*) values calculated from rdc data for all relative orientations of the C-terminal domain with respect to the N-terminal domain (in blue) of CaM in the presence of AS. (B) Conformations with the largest MAP values in agreement with both rdc and pcs data. Details same as those for Figure 6.

the C-terminal domain (the vector connecting the atoms C′ of residue Ala 88 and C^{α} of residue Gly 113 is a suitable one) and the direction of the North pole (just like the direction provided by the needle of a compass on the surface of Earth; see Figure 6D). This angle has been selected because of its property to monitor the rotation of the C-terminal domain around its first helix independently of the orientation of the latter.

In the case of free CaM, five regions have very similar MAP(*R*) values, equal to 0.37−0.386. In all of them (Figure 6A) the first helix of the C-terminal domain forms quite large angles with the last helix of the N-terminal domain, and in four orientations it is directed parallel to the direction of the β-sheet present in the N-terminal domain.

Minimizations were then performed using pcs and rdc data, for increasing the weight of a few selected orientations, as shown in Figure 8A. Translations were restrained so that the distance between the last C^{α} atom of the N-terminal domain (C^{α} of Asp 78) and the first atom of the C-terminal domain (C^{α} of Ser 81) cannot exceed that given by the fully extended conformation of the intervening residues (i.e., it is not larger than 9 Å). The absolute minimum value of the target function allowed by the experimental data was 0.215, and the threshold ϵ for admissible solutions was set to a 10% larger value, i.e., 0.236. The starting orientations to be provided to the minimization program were selected within the Euler angle space representing 20% of the orientations with the largest MAP(*R*) values. The largest weights of these orientations allowing a target function smaller than ϵ were used to rescale the MAP(*R*) values. The results are shown in Figure 6B, where only the conformations with largest MAP values in agreement with both pcs's and rdc's are depicted. The largest MAP value was found to be 0.365, and the corresponding

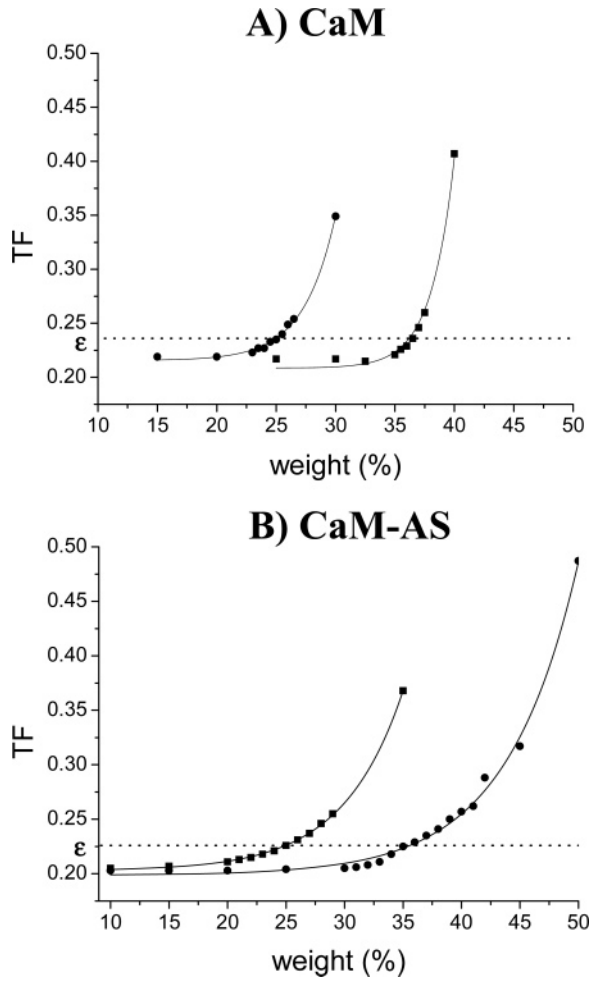


Figure 8. Target function $TF(w_0)$ for two different orientations in bad or good agreement with the experimental data for the CaM (A) and CaM−AS (B) cases. The TF function has a roughly exponential behavior, as shown by the fits. The maximum weight corresponding to a *TF* value equal to ϵ (shown as dotted line) defines the MAP for such a conformation.

conformations are relative to the few orientations labeled with "OR1" in Figure 6A. This clearly shows that such orientations are in best agreement with the pcs data. Conformations corresponding to orientations labeled with "OR2", "OR4", and "OR5" in Figure 6A may be heavily represented, because the MAP for such conformations is calculated to be around 0.35 (see Figure 6B). It cannot be excluded that some of these are actually ghosts. Conformations corresponding to orientations labeled with "OR3" are less preferred, as their MAP decreases to 0.33. These results represent a significant refinement of those reported in Bertini et al.[36]

In the case of CaM−AS the preferred orientations have somewhat larger MAP(R) values than those for free CaM (see Figure 5), as a result of the larger rdc's measured for the C-terminal domain nuclear pairs. The four preferred orientations have MAP(R) values up to 0.39−0.434. Some orientations are similar to those obtained for free CaM, but one orientation is also present in a region near that of the closed conformation of the protein (Figure 7).

The absolute minimum for the target function allowed by the experimental pcs and rdc data was 0.205, and ϵ was fixed to 0.226 (Figure 8B). The largest MAP was found to be 0.35, and the corresponding conformations obtained from the minimization program are those reported in Figure 7B. In such structures, the first helix of the C-terminal domain is tilted to about $\alpha \approx 110°$ with respect to the direction of the last helix of the N-terminal domain and forms an angle of about $\beta \approx 90°$ with respect to the plane containing the axis of the last helix of the N-terminal domain and the calcium ion in its second binding loop. These conformations correspond to the orientations labeled with "OR3" in Figure 7A. The solutions obtained indicate that, in the presence of AS, the conformations of CaM with the largest MAP values are not far from the closed conformation ($\alpha \approx 110°$, $\beta \approx 100°$) observed in PDB structures 1PRW and 2BBM. All other orientations, i.e., those labeled with "OR1", "OR2", and "OR4" in Figure 7A are in worse agreement with the pcs data. The MAP value of the conformation derived from "OR1" is only 0.29, and the other two are even lower. Therefore, the use of pcs's allowed us to efficiently rank the conformations with the largest MAP values.

In conclusion, the above analysis indicates that free CaM adopts a large ensemble of conformations, none of them with an MAP larger than 0.36, which are quite different from the closed conformation, in agreement with results reported by Bertini et al.[36] In the adduct with AS, CaM still adopts a large ensemble of conformations, but in this case the conformations with the largest MAP values are in a region of space close to that occupied by the closed conformation, with an MAP not larger than 0.35.

Concluding Remarks and Perspectives

A novel method has been proposed for the structural characterization of systems displaying conformational heterogeneity, constituted by substructures considered rigid and relatively free to move with respect to each other. Such substructures may be interacting proteins not rigidly connected or different protein domains within the same protein. The method is generally valid and can be applied whenever a paramagnetic ion is attached to one substructure, and the effects are observed in the other substructure(s).

For the first time a quantitative assessment of the conformational space experienced by a protein consisting of two domains relatively free to move with respect to each other is provided in terms of the maximum allowed probability (MAP) for each conformation. The procedure is rigorous in setting an upper limit to the percent occupation of a given conformation. In this sense, ghosts are not a problem; they only make some nonpreferred conformations less nonpreferred. In no case can a conformation be in reality more allowed than calculated.

The results are not only consistent with our previous analysis[36] performed on CaM but also more solid, thanks to a more rigorous mathematical treatment. In the CaM−AS adduct, the conformations with the largest MAP values experienced by calmodulin are reminiscent of those observed for the same protein when interacting with peptides with high affinity.

The power of the method is expected to increase with decreasing conformational freedom, as long as conformational heterogeneity is still present to some extent. In fact, systems experiencing less conformational freedom have larger averaged rdc and pcs values, which means less percent error and lessghosts. This results in higher accuracy in the identification of the conformational space experienced. Free CaM is, in this

respect, a difficult case, and yet the method works reasonably well. For CaM−AS it works better. In a case where there is more limited (but still relevant) conformational freedom, the method would be maximally powerful.

Acknowledgment. This work has been supported by Ente Cassa di Risparmio, Ministero dell'Università e della Ricerca COFIN 2005, and by the European Commission, Contract LSHG-CT-2006-031220 (SPINE II), Contract EU-NMR 026145 (JRA 3 ORIENTING-NMR), Contract LSHG-CT-2004-512052 (UPMAN), and Contract LSHG-CT-2004-512077 (NDDP). Cristina del Bianco performed an early set of NMR measurements on the CaM−AS interaction.

Supporting Information Available: Preliminary considerations on pcs and rdc values; the algorithm; calculation of MAP-(R) from rdc data; calculation of MAP from pcs and rdc data; synthetic tests; preparation of AS. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0726613