University of Florence

International Doctorate in Structural Biology

Cycle XXI (2006-2008)

Structural and dynamic characterization of Ca^{2+} binding proteins by NMR spectroscopy

PhD thesis of

Jing Yuan

Tutor	Coordinator
Prof. Giacomo Parigi	**Prof. Claudio Luchinat**

S.S.D. CHIM/03

This thesis has been approved by the University of Florence,

the University of Frankfurt and the Utrecht University

Content

1
INTRODUCTION

1.1 Ca^{2+} in biological systems

Calcium is an abundant "inorganic element" in biological systems. Of the approximately 1400 g of calcium in the human body, most of it is immobilized in bones and teeth as hydroxyapatite.[1] A minor part circulates as free Ca^{2+} in the blood and in the extracellular space or is stored intracellularly in distinct compartments.[2]

Ca^{2+} ions, as recognized widely, are central to a complex intracellular messenger system that is mediating a wide range of biological processes, i.e. muscle contraction, secretion, glycolysis and gluconeogenesis, ion transport and cell division. The concentration of Ca^{2+} is largely ranging across the plasma membrane, where extracellular concentrations are 10,000 times higher than intracellular ones. External signals, such as hormones, light, stress or pathogenesis, can often lead to transient increases in calcium concentrations within the cell.[1] The increasing Ca^{2+} concentration is the result of either the influx of extracellular Ca^{2+} or the release of this cation from internal stores, particularly the ER/SR (endoplasmic/sarcoplasmic reticulum), which lead to calcium binding by regulatory proteins. The influx of Ca^{2+} ions contributes the switch of many cellular processes, turning them from an ‘off’ state to an ‘on’ state.[3-5] For the reverse functional switch, Ca^{2+} ions are actively pumped outside the cell across the membrane by the plasma membrane calcium ATPase (PMCA pump).[6] The concentrations of free cellular Ca^{2+}, being in ‘off’ state, are maintained at the level of 100–200 nM.

Ca^{2+} ions are also known to play various roles outside cells. In the blood plasma of mammals, in which the Ca^{2+} concentration exceeds the intracellular by a factor of about 10^4, Ca^{2+} ions are instrumental in joining certain proteins in the blood-clotting system with membrane surfaces of circulating cells. Many extracellular enzymes also contain Ca^{2+} ions, sometimes at the active site but most often at other locations. It is generally believed that Ca^{2+} ions confer an increased thermal stability on proteins, and indeed proteins in heat-tolerant microorganisms often hold many Ca^{2+} ions.[7,8]

1.2 Ca^{2+} -binding proteins

When Ca^{2+} flows into the cytoplasm during the ‘on’ state, it becomes bound to a wide variety of Ca^{2+}-binding proteins, many of which belong to a homologous family defined by

helix–loop–helix secondary structure termed the 'EF-hand' motif.[9] Functionally, EF-hand proteins can roughly be divided into two general classes: the Ca^{2+} sensors and the Ca^{2+} buffers. The Ca^{2+} sensors translate the chemical signal of an increased Ca^{2+} concentration into diverse biochemical responses. This signal transduction is accomplished predominantly through a Ca^{2+}-induced conformational change, as illustrated by the classic examples of calmodulin (CaM), recoverin or S100 proteins. The Ca^{2+} buffers are a smaller subset of the EF-hand protein family. Exemplified by calbindin D9k and parvalbumin, these proteins help to modulate the Ca^{2+} signal both spatially and temporally as they bind the free Ca^{2+} to transmit the signal throughout the cell or to remove the potentially harmful ion from the cytoplasm.[10,11]

This classical EF-hand motif is characterized by a sequence of 12 amino acid residues with the pattern X*Y*Z*–Y*–X**–Z. (Figure 1) The positions X, Y, Z, –X, –Y and –Z represent the ligands, which participate in metal coordination, and the stars represent the intervening residues. Strong preferences exist for aspartate and glutamate in the 1 and 12 coordinating positions respectively. The sixth residue in the loop is necessarily glycine due to the conformational requirements of the backbone. The intervening residues are typically hydrophobic and form a hydrophobic core that stabilizes the two helices. Functional EF-hands are found in pairs and are required for the correct folding of the proteins and unique variations of calcium binding co-operativity[12].

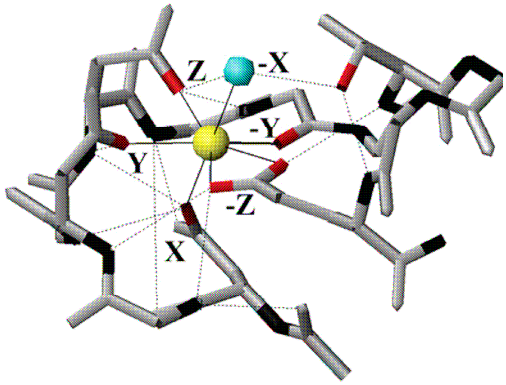


Figure 1: Ca^{2+} co-ordination by the canonical EF-hand illustrating both the pentagonal bipyramidal co-ordination of the Ca^{2+} ion (continuous lines) and the extensive hydrogen bonding pattern found in the loop (broken lines). The backbone NH groups are indicated in black, the side-chain oxygen atoms in red, the Ca^{2+} ion in yellow and the co-ordinating water in blue (PDB code 1EXR) [13] (reprint from Gifford, J.L., Biochem. J. (2007) 405, 199–221);

1.3 CaM and its cellular function

Calmodulin (CaM) is a ubiquitous, Ca^{2+} -binding protein, which can bind to and regulate a multitude of different protein targets, thereby affecting many different cellular functions, i.e. metabolism, cytoskeletal dynamics, cell proliferation, cell–cell interaction and development.[14] Many of the proteins that CaM binds are unable to bind calcium themselves, and as such use CaM as a calcium sensor and signal transducer. These phenomena have been mediated by not only the Ca^{2+}-CaM complex, but also by apo-CaM. After intracellular Ca^{2+} concentration increases, lead by the activity of N-methyl-d-aspartate (NMDA) receptors or voltage-sensitive Ca2+ channels, CaM releases from neuromodulin or neurogranin, where CaM was previously bound. Depending on Ca^{2+}, CaM interacts with at least 30 different enzymes and proteins that modulate the activity of several key signaling molecules, including adenylyl cyclases (AC), protein kinases, calcineurin, nitric oxide synthase, Ca^{2+}-channels, ATP-dependent Ca^{2+}-pumps, and the CaM-dependent protein kinases (CaMKII) (Figure 2). ApoCaM also binds to a distinct set of proteins such as actin-binding proteins, cytoskeletal and membrane proteins, enzymes, receptors and ion channels.[14,15] In fact, most CaM in unstimulated cells (‘off’ state) would exist in the state of apoCaM because CaM level in eukaryotic cell is approximately 1–10 μM while free calcium concentration in unstimulated cell is 0.1 μM.[15]

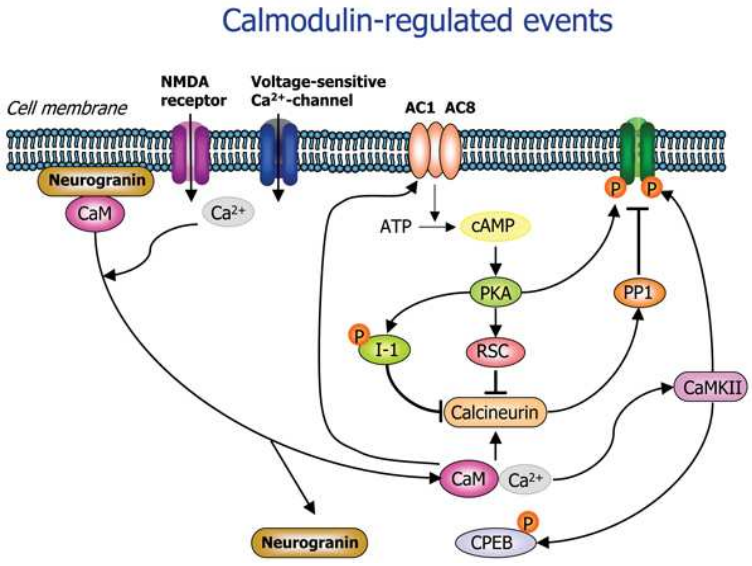


Figure2: Ca^{2+}-CaM network. (reprint from Klipp and Liebermeister BMC Neuroscience 2006 7(Suppl 1):S10)

In the result section of this thesis, two CaM binding proteins were investigated:

1) α-synuclein,

α-Synuclein is a 140 amino acid protein abundantly expressed in presynaptic terminals

of vertebrates and plays the central role in Parkinson's disease and other diseases involving Lewy bodies. Subcellular fractionation experiments suggested that there might be a certain exchange mechanism of α-synuclein between membranes and cytosol.[16] Therefore, CaM interaction of α-synuclein respected to examine the possible implication of α-synuclein in the calcium-dependent process of vesicular biogenesis.[17] The dissociated constant of the complex of α-synuclein and CaM is in μM range.[18] It might be speculated that α-synuclein could act in cellular signaling processes by interacting with CaM during the mediation of calcium signaling.

2) CaM binding with Death-associated protein kinase

Death-associated protein kinase (DAPk) is a serine/threonine protein kinase, regulated by CaM and is a drug-discovery target for neurodegenerative disease.[19] *In vivo* experiments showed that injury of inhibitor of DAPk to animal could enhance neuronal survival and reduce brain tissue loss, suggesting that DAPk has a functional role in mediating death pathways in neurodegenerative disorders.[20] However, the pathway in the DAPk-mediated signal transduction is still mystery. A proposed mechanism for DAPk regulation is shown in Figure 3A. Upon the apoptotic stimulus, the functional autophosphorylation site is dephosphorylated, leading to DAPk activation. The CaM binding site (green in the figure) thus is able to bind CaM as a result of 'weakened lock' and the enzyme facilitates full activation.[21] Four additional kinases have been identified based on the high homology of their catalytic domain to that of DAPk. One member of this family, DRP-1, also can be activated by Ca^{2+}-CaM. DRP-1 activation in apoptosis provides different features, i.e. that of forming homodimers after dephosphorlation (Figure 3B).[21]

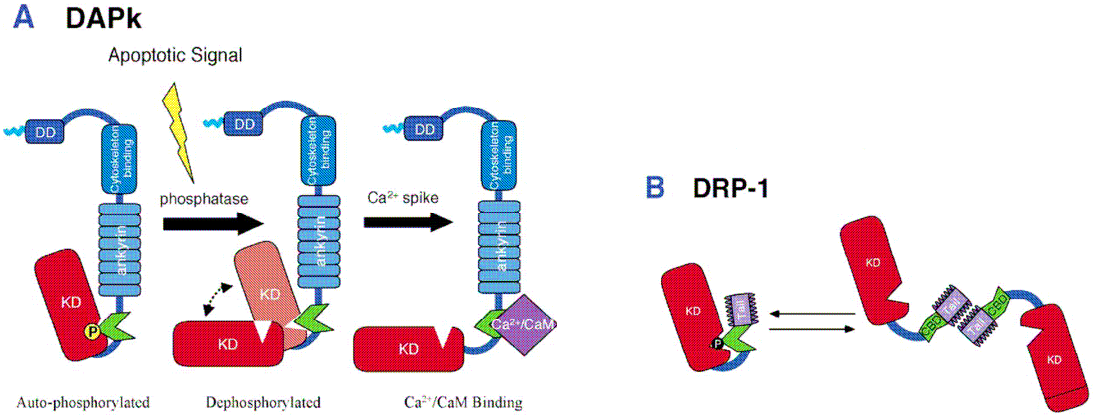
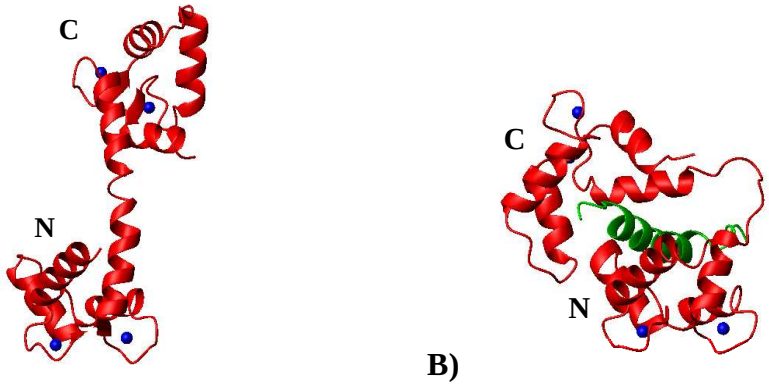


Figure 3.Specific features that characterize the mode of DAPk and DRP-1 activation. The various domains are marked: KD (kinase domain); DD (death domain). (Reprint form Shohat G, Biochimica et Biophysica Acta 1600 (2002) 45– 50)

1.4 Structure features and inter-domain motion of CaM

CaM is a two-domain protein, containing 148 amino acid residues. Each domain is made up of two EF-hands motifs and binds two Ca^{2+} ions. The two domains are connected by a short linker. Early X-ray data showed that CaM have a dumbbell shape, with helix 4, the last helix of the N-terminal domain, and helix 5, the first helix of the C-terminal domain, together with the interdomain linker, forming a long continuous helical structure.[22] (Figure 4. A) It was soon recognized that the NMR properties of CaM in solution were inconsistent with the rigid dumbbell shape observed in the early X-ray work, and that the central part of the helix loses its helical character and allows reciprocal reorientation of the two domains.[23-26] An extended model-free analysis characterized the relative motions as occurring on a time scale of about 3 ns. [24]

Upon forming complexes with its targets, the interdomain motion of the two domains of CaM may be lost. For example, the first structure of Ca^{2+}-CaM binding with a skMLCK peptide, was solved by NMR in 1992.[27] (Figure 4. B) As shown from that structure, the N-terminal and the C-terminal domains of CaM wrap around the bound peptide, which has an α-helical structure, according to the canonical closed state. The skMLCK peptide binds in an antiparallel orientation, i.e. with the N-terminal and the C-terminal CaM domains interacting mainly with the C-terminal and N-terminal halves of the peptide, respectively. The interdomain motion observed in the CaM free form, is not present any more, as indicated by NMR studies. The losing of this interdomain motion resides on the fact that the hydrophobic residues of the peptide anchor their side chains to the solvent exposed hydrophobic residues of both domains of CaM, forming stable interactions, and thus freezing the CaM into a single conformation.[28-31] Similar binding mode was also detected for the CaM complexes with the endothelial nitric oxide synthase peptide, and with a peptide derived from the olfactory CNG channel.[27] Those peptides share a common, so called, 1-14 motif, which features two hydrophobic anchor residues located at first and fourteen spacings in the primary sequence. Other binding models have also been identified, termed 1-10, 1-16 or 1-17 motifs based on the position of the two key anchoring hydrophobic residues in the target peptide.[27]



A) B)

Figure 4: Relative orientation of the N- and C-terminal domains in CaM as observed by X-ray in the absence of target peptides (A; extended conformation, PDB code: 3cln) and as observed in the presence of target peptides (B; closed conformation, PDB code: 2bbm)

The interdomain motion of CaM observed when the protein is free in the solution, is maintained in CaM-complexes, when the peptide partners only interact with one domain of CaM. For example, the plasma membrane calcium pump peptide, named C20W [32], is only anchored on the C-domain of CaM, thus allowing the N-terminal to move freely with respect to the C-domain. We have shown that this property is also been observed when CaM binds α-synuclein. In some unusual cases, some order of freedom may remain even if both domains of CaM interact with a target. In 2006, the crystal structure of Ca^{2+}CaM bound to skeletal muscle ryanodine receptor (RYR1) peptide reveals that CaM recognizes two hydrophobic anchor residues at a '1-17' motif, but residual dipolar couplings show that the two domains of CaM experience domain motions within the complex. [33]

Conformational flexibility is a crucial feature in the mechanism of action of a number of proteins/enzymes.[34-36] In essence, the basic tools are still lacking 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. NMR relaxation parameters have already been used to detect local and global motion of CaM.[37,38] However, unfortunately, this technique can not provide details on the probability of a given conformation, when the interdomain motion is present in the system.

In rigid complexes between CaM and its partners, another problem may exist. The inter molecular packing force, which occurs frequently in the solid state structure, may easily alter

the relevant orientation of the two domains of CaM, making the conformation different with respect to that one in solution, even when the peptide is anchored to both domains.[39-41] NMR spectroscopy has the advantage of providing the conformation of protein in solution, but is limited by relative low precision, the best NMR structures being obtained with a resolution corresponding to that of about 2Å resolution for a crystal structure.[42] So it is still an open question to get structure information and evaluate the accuracy of the structure of CaM bound to its targets. We have proved that inclusion of few selected NMR restraints can increase the accuracy in solution of a high-resolution structural model, which was provided by the X-ray structure.

1.5 Structure features of S100 protein

S100 proteins are another Ca^{2+} binding, EF-hand protein family, comprising 20 known human members each coded by a separate gene. At least 16 of these genes cluster to chromosome 1q21, known as the epidermal differentiation complex. The S100 proteins are small, acidic proteins of 10-12 kDa, found exclusively in vertebrates[43]. Members of this protein family have been implicated in the Ca^{2+}-dependent (and, in some cases, Zn^{2+} or Cu^{2+} dependent) regulation of a variety of intracellular activities. Intracellular functions include regulation of protein phosphorylation and enzyme activity, calcium homeostasis, regulation of cytoskeletal components and regulation of transcriptional factors [44] (Figure 5). A number of S100 proteins interact with p53, however, they exert different effects on p53 activity. Both S100A4 and S100B are thought to inhibit p53 phosphorylation leading to inhibition of its transcriptional activity, thereby compromising p53 tumour-suppressor activity. In contrast S100A2 promotes p53 transcriptional activity and interestingly S100A4 has also been documented as enhancing p53-dependent apoptosis. Thus the balance of actions of different S100 proteins within a cell will determine function.[45-47] Many of the S100 family members have a role in modulating cytoskeletal dynamics, i.e. both S100A1 and S100A11 have been shown to inhibit cell proliferation.[45] Some S100 members can be released or secreted into the extracellular space. When they are present extracellularly, some members of the family act as leukocyte chemoattractants, macrophage activators and modulators of cell proliferation.[48] These functions associate S100 proteins with a variety of pathologies such as inflammation

and cardiomyopathies, and gradually their role in carcinogenesis is beginning to unravel. It is clear that some S100 proteins act as tumour promoters and others as tumour suppressors, and there is exquisite tissue specificity in their actions.

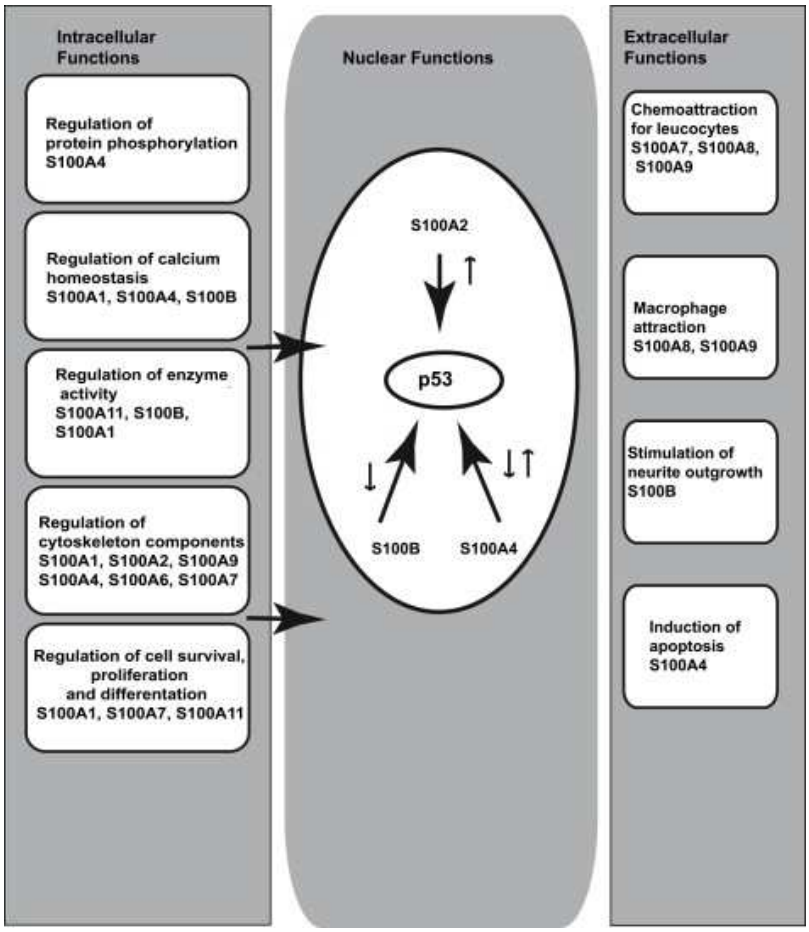


Figure 5: S100 protein and their function (Reprint from I. Salama et al. Eur J Surg Oncol.2008Apr; 34(4): 357-64)

Structural data showed that many S100 members exist within cells as homodimers, heterodimers and oligomers. Generally, S100 proteins are organized as tight homodimers, in which the two monomers are related by a two-fold axis of rotation. Each monomer has two distinct EF-hands, one is a C-terminal, 'canonical' EF-hand, and common to all EF-hand proteins, and the other is a N-terminal, 'pseudo' EF-hand, characteristic of S100 proteins.[44] Upon Ca^{2+}-binding, the C-terminal EF-hand undergoes a large conformational change resulting in the exposure of a hydrophobic surface responsible for target binding, indicating their specific biological properties. (Figure 6)

S100A5 is a novel member of the EF-hand superfamily of calcium-binding proteins that is poorly characterized at the protein level. Immunohistochemical analysis demonstrates that

it is expressed in very restricted regions of the adult brain[49]. Flow dialysis revealed that the homodimeric S100A5 binds four Ca^{2+} ions with strong positive cooperativity and an affinity 20–100-fold higher than the other S100 proteins studied under identical conditions. S100A5 also binds two Zn^{2+} ions and four Cu^{2+} ions per dimer.[50] None of these ions change the α-helical-rich secondary structure.[50] Although the structures and functional role of some S100 proteins have been characterized, no structural data was available for both the apo and the Ca^{2+} form of S100A5. Also there are only few published works on the dynamic properties of S100 protein, which could provide important features for ligand binding. Furthermore, the Zn^{2+} binding sites for different S100 proteins are all different with each other. The Zn^{2+} binding sites of S100A5 are still unknown.

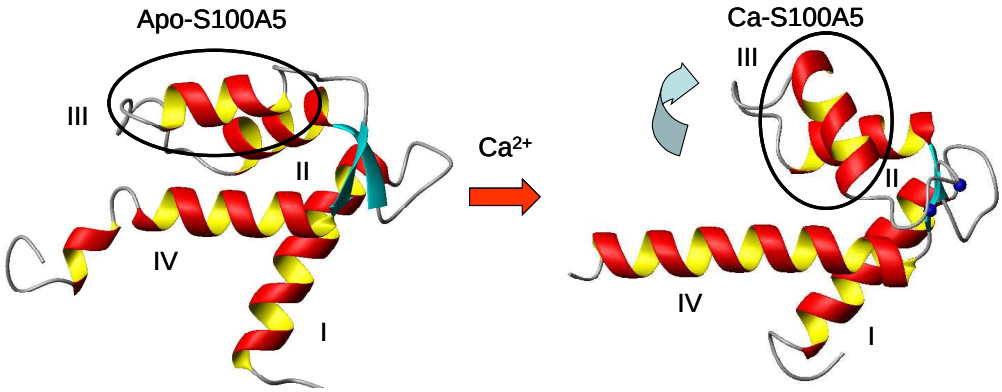


Figure 6 Major structural differences upon calcium binding between apo and Ca-S100A5

1.6 Aims and topics of the research

During the three years of the Ph.D. course, my research was focused on the investigation of the structural and dynamic characterizations of Ca^{2+} binding protein by NMR spectroscopy.

As already described in section 1.4, CaM is a two-domain protein and able to bind target peptides by wrapping the two domains around the target; such domains are free to move with respect to one another in the peptide-free from. Novel approaches are needed to monitor the relative positions that can be experienced by the two domains, the relative weight of each conformation and time scale of the motions involved. Both pcs and rdc are long-range restraints and, therefore, optimally suited to detect global structural features, such as relative orientations of secondary structure elements or entire domains. Pcs and rdc are also time average values when the motions occur on time scales faster than ms. Thus, in principle, pcs and rdc data contain information on the probability of any conformation that can be

experience by the proteins connected by multiple domains.

The aim of my research was to develop new approaches to understand the structural and dynamic properties of multidomain proteins by using experimental pcs and rdc data, originated from paramagnetism. Such approach was applied for a conformational characterization of CaM, when free in solution, in the presence of the target protein α-synclein and in the presence of other peptides. The work included: i) judging the presence of interdomain motion by rdc data, ii) determining the maximum allowed probabilities of any given conformation when the interdomain motion is present, and iii) refining crystal structural models to obtain high quality and precision solution structure if the two domains are fixed. In the last case, packing force of crystal structure can be corrected.

My research work also included the structure, dynamics and metal binding characterization of the S100A5 protein. S100A5 is expressed in the region of the adult brain and only little work has been performed to understand the function of this protein. The solution structures of S100A5 in both the apo and the calcium forms have been studied, and relaxation studies have been performed to monitor its mobility. Its binding affinities with other metal ions, like Zn^{2+}, are also investigated.

Reference List

(1) Carafoli, E. *Proc.Natl.Acad.Sci.USA* **2002**, *99*, 1115-1122.

(2) Carafoli, E. *Nature Reviews Molecular Cell Biology* **2003**, *4* , 326-332.

(3) Berridge, M. J. *Bioessays* **1995**, *17*, 491-500.

(4) Berridge, M. J. *Novartis Found Symp* **2001**, 52-64.

(5) Berridge, M. J.; Bootman, M. D.; Roderick, H. L. *Nat Rev Mol Cell Biol* **2003**, *4*, 517-529.

(6) Di Leva, F.; Domi, T.; Fedrizzi, L.; Lim, D.; Carafoli, E. *Archives of Biochemistry and Biophysics* **2008**, *476*, 65-74.

(7) Wiest, M. C.; Eagleman, D. M.; King, R. D.; Montague, P. R. *Journal of Neurophysiology* **2000**, *83*, 1329-1337.

(8) Egelman, D. M.; Montague, P. R. *Biophysical Journal* **1999**, *76*, 1856-1867.

(9) Baimbridge, K. G.; Celio, M. R.; Rogers, G. H. *Trends Neurosci* **1992**, *15*, 303-308.

(10) Grabarek, Z. *Journal of Molecular Biology* **2006**, *359*, 509-525.

(11) Capozzi, F.; Casadei, F.; Luchinat, C. *J.Biol.Inorg.Chem.* **2006**, *11*, 949-962.

(12) Strynadka, N. C. J.; James, M. N. G. *Annu.Rev.Biochem.* **1989**, *58*, 951-998.

(13) Wilson, M. A.; Brunger, A. T. *J.Mol.Biol.* **2000**, *301*, 1237-1256.

(14) Zhang, M. J.; Yuan, T. *Biochemistry and Cell Biology-Biochimie et Biologie Cellulaire* **1998**, *76*, 313-323.

(15) Jurado, L. A.; Chockalingam, P. S.; Jarrett, H. W. *Physiological Reviews* **1999**, *79*, 661-682.

(16) Davidson, W. S.; Jonas, A.; Clayton, D. F.; George, J. M. *Journal of Biological Chemistry* **1998**, *273*, 9443-9449.

(17) Oneil, K. T.; Degrado, W. F. *Trends in Biochemical Sciences* **1990**, *15*, 59-64.

(18) Slemmon, J. R.; Morgan, J. I.; Fullerton, S. M.; Danho, W.; Hilbush, B. S.; Wengenack, T. M. *Journal of Biological Chemistry* **1996**, *271*, 15911-15917.

(19) Cohen, O.; Feinstein, E.; Kimchi, A. *The EMBO Journal* **1997**, *16*, 998-1008.

(20) Schumacher, A. M.; Velentza, A. V.; Mirzoeva, S.; Watterson, D. M. *Faseb Journal* **2002**, *16*, A581.

(21) Shohat, G.; Shani, G.; Eisenstein, M.; Kimchi, A. *Biochimica et Biophysica Acta-Proteins and Proteomics* **2002**, *1600*, 45-50.

(22) Babu, Y. S.; Bugg, C. E.; Cook, W. J. *J.Mol.Biol.* **1988**, *204*, 191-204.

(23) Barbato, G.; Ikura, M.; Kay, L. E.; Pastor, R. W.; Bax, A. *Biochemistry* **1992**, *31*, 5269-5278.

(24) Baber, J. L.; Szabo, A.; Tjandra, N. *J.Am.Chem.Soc.* **2001**, *123*, 3953-3959.

(25) Seaton, B. A.; Head J.F.; Richardson, F. M. *Biochemistry* **1985**, *24*, 6740-6743.

(26) Heidorn, D. B.; Trewhella, J. *Biochemistry* **1988**, *27*, 909-915.

(27) Ikura, M.; Clore, G. M.; Gronenborn, A. M.; Zhu, G.; Clee, C.; Bax, A. *Science* **1992** , *256*, 632-638.

(28) Schumacher, M. A.; Rivard, A. F.; Bächinger, H. P.; Adelman, J. P. *Nature* **2001**, *410*, 1120-1124.

(29) Ikura, M.; Barbato, G.; Klee, C. B.; Bax, A. *Cell Calcium* **1992**, *13*, 391-400.

(30) Meador, W. E.; Means, A. R.; Quiocho, F. A. *Science* **1992**, *257*, 1251-1255.

(31) Osawa, M.; Tokumitsu, H.; Swindells, M.; Kurihara, H.; Orita, M.; Shibanuma, T.; Furuya, T.; Ikura, M. *Nat Struct Biol* **1999**, *6*, 819-826.

(32) Elshorst, B.; Hennig, M.; Forsterling, H.; Diener, A.; Maurer, M.; Schulte, P.; Schwalbe, H.; Griesinger, C.; Krebs, J. F.; Schmid, H.; Carafoli, E. *Biochemistry* **1999**, *38*, 12320-12332.

(33) Maximciuc, A. A.; Putkey, J. A.; Shamoo, Y.; MacKenzie, K. R. *Structure* **2006**, *14*, 1547-1556.

(34) Huang, Y. J.; Montelione GT *Nature* **2005**, *438*, 36-37.

(35) Ishima, R.; Torchia, D. A. *Nature Struct.Biol.* **2000**, *7*, 740-743.

(36) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128-132.

(37) Mittermaier, A.; Kay, L. E. *Science* **2006**, *312*, 224-228.

(38) Fischer, M. W. F.; Zeng, L.; Majumdar, A.; Zuiderweg, E. R. P. *Proc.Natl.Acad.Sci.USA* **1998**, *95*, 8016-8019.

(39) Chou, J. J.; Li, S.; Klee, C. B.; Bax, A. *Nature Struct.Biol.* **2001**, *8*, 990-997.

(40) Skrynnikov, N. R.; Goto, N. K.; Yang, D.; Choy, W.-Y.; Tolman, J. R.; Mueller, G. A.; Kay, L. E. *J.Mol.Biol.* **2000**, *295*, 1265-1273.

(41) Goto, N. K.; Skrynnikov, N. R.; Dahlquist, F. W.; Kay, L. E. *J.Mol.Biol.* **2001**, *308*, 745-764.

(42) Fragai, M.; Luchinat, C.; Parigi, G. *Acc.Chem.Res.* **2006**, *39*, 909-917.

(43) Heizmann, C. W.; Fritz, G.; Schafer, B. W. *Frontiers in Bioscience* **2002**, *7*, D1356-D1368.

(44) Donato, R. *Int.J.Biochem.Cell Biol.* **2001**, *33*, 637-668.

(45) Donato, R. *Microsc.Res.Tech.* **2003**, *60*, 540-551.

(46) Mueller, A.; Schafer, B. W.; Ferrari, S.; Weibel, M.; Makek, M.; Hochli, M.; Heizmann, C. W. *Journal of Biological Chemistry* **2005**, *280*, 29186-29193.

(47) Grigorian, M.; Andresen, S.; Tulchinsky, E.; Kriajevska, M.; Carlberg, C.; Kruse, C.; Cohn, M.; Ambartsumian, N.; Christensen, A.; Selivanova, G.; Lukanidin, E. *Journal of Biological Chemistry* **2001**, *276*, 22699-22708.

(48) Hiratsuka, S.; Watanabe, A.; Aburatani, H.; Maru, Y. *Nature Cell Biology* **2006**, *8*, 1369-1U31.

(49) Wicki, R.; Schafer, B. W.; Erne, P.; Heizmann, C. W. *Biochem Biophys Res Commun* **1996**, *227*, 594-599.

(50) Schafer, B. W.; Fritschy, J. M.; Murmann, P.; Troxler, H.; Durussel, I.; Heizmann, C. W.; Cox, J. A. *Journal of Biological Chemistry* **2000**, *275*, 30623-30630.

2
METHODOLOGICAL ASPECT

Nuclear magnetic resonance (NMR) spectroscopy is a bio-physical method which can provide high-resolution structures of biological molecules such as proteins and nucleic acids and their complexes at atomic resolution. In the NMR experiments, solution conditions such as the temperature, pH and salt concentration can be adjusted so as to closely mimic a given physiological fluid. Importantly, biomolecular NMR spectrocopy can provide information about conformational dynamics and exchange processes of biomolecules at timescales ranging from picoseconds to seconds, and is very efficient in determining ligand binding, mapping interaction surfaces and studies thermodynamic and kinetic aspects of interactions of protein/ligand complexes.[1-4] Recently, paramagnetic restraints, such as pseudocontact shifts (pcs) and residual dipolar couplings (rdc) have been used as long-range structural restrains in order to improve both precision and accuracy of structure determination in solution. Besides additional restraints during structure calculation, pcs and rdc data are extremely useful for defining the relative orientation for domains of multi-domain proteins and describing inter-domains dynamics, if present in multi-domain protein.[5,6] Here, I focus on the methodologies used for 1) the structure determination and dynamic studies on the S100A5 protein; 2) monitoring the relative orientation of CaM by using paramagnetic constraints; 3) refining the X-ray structure of the CaM-peptide complex by using paramagnetic constraints.

2.1 Structure determination by NMR spectroscopy

The standard protocol for NMR structure determination includes the preparation of a homogeneous protein solution, the recording and handling of the NMR datasets, the structural interpretation of the NMR data. Typically, uniformly ^{13}C and ^{15}N labelled proteins are used for structure determination. The protocol of resonance assignments of double (^{15}N and ^{13}C) labelled proteins using 3D experiments is based on a number of experiments showing cross peaks among the backbone or the side chain nuclei. The details are described elsewhere.[7]

Traditionally, the most important structural information derived from NMR spectra is based on the nuclear Overhauser effect (NOE), which is a result of cross-peak between different spins (normally protons) in a molecule and depends on the through-space distance between these spins. NOEs are typically only observed between protons which are separated by less than 5-6 Å. Because the intensity of the cross peaks are also affected by motions of the

molecule, NOE data are usually treated as upper bounds (UPL) on inter atomic distances instead of precise distance restraints. For the same reason, the absence of an NOE is in general not interpreted as a lower bound on the distance between the two interacting spins.

The structure quality will be much improved, if calculations are performed by adding torsion angle restraints to distance restraints. Several programs are available for the prediction of the secondary structure elements, such as TALOS and PREDICTOR, by using a complete set of backbone chemical shifts for all H_α, C_α, C_β and CO resonances. In this thesis, TALOS program was used to predict torsion angle restraints.[8] In solving the solution structure of S100A5, NOE and angle constrains were used.

During the years, new constraints have been used into structure calculations: angles obtained from cross-correlation effects,[9] residual dipolar couplings (rdc) [10-13] and hydrogen bonds.[14] The strong interest for new constraints arises from the need of solving structures without NOEs or with relatively few of them. In proteins that are able to specifically bind paramagnetic metal ions, paramagnetism-based constraints can be exploited.[15] They are the contact shifts, pseudocontact shifts (pcs), the hyperfine shifts as sum of the two, residual dipolar couplings (rdc) and paramagnetic relaxation enhancement effect (pre).[15,16, 21,22] Paramagnetism-based pcs and rdc have been, in particular, used for the structural characterization of CaM.

2.2 Dynamic properties characterization by NMR relaxation parameters

Protein dynamics over a wide range of time scales and amplitudes plays an important role in biological functions, such as enzyme reaction, ligand binding, and folding. NMR relaxation parameter technique provides unique opportunities to explore a wide range of dynamic time scales such as fast internal motion (ps-ns), slow motion (μs-ms), the overall correlation time of a protein (of the order of 10 ns).[2,17,18]

The majority of applications of spin relaxation methods in proteins utilize the amide ^{15}N spin as a probe of backbone motions. The dominant relaxation mechanisms are the magnetic dipolar (DD) and anisotropic chemical shift (CSA) mechanisms.[19] Equations describing relaxation parameters in terms of spectral density functions are given by [20]:

$$R_1 = \frac{D^2}{4}[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + \frac{C^2}{3} J(\omega_N) \quad \textbf{(1)}$$

$$R_2 = \frac{D^2}{8}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)]$$
$$+ \frac{C^2}{18}[4J(0) + 3J(\omega_N)] + R_{ex} \quad \textbf{(2)}$$

$$\mathrm{NOE} = 1 + (\gamma_N / \gamma_H)\frac{D^2}{4}[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]/R_1 \quad \textbf{(3)}$$

Where

$$D = \frac{\mu_0 h \gamma_N \gamma_H}{8\pi^2} < r_{HN}^{-3} > \quad \textbf{(4)}$$

$$C = \omega_N \Delta\sigma \quad \textbf{(5)}$$

$$J(\omega) = \frac{2}{5}[\frac{\tau_c}{1 + \omega^2 \tau_c^2}] \quad \textbf{(6)}$$

In eqs. (4) and (5), μ_0 = 4π × 10^{-7} kg m s^{-2} A^{-2}, h = 6.6262 × 10^{-34} erg · s, where h is Planck's constant; γ_H and γ_N are the gyromagnetic ratios of ^{1}H and ^{15}N respectively (2.6753 × 10^8 rad · s^{-1} · T^{-1} and -2.71 × 10^7 rad · s^{-1} · T^{-1}); ω_H and ω_N are the Larmor frequencies; Δσ is the anisotropy of the chemical shift tensor of the ^{15}N spin (160ppm or 170ppm); r_{HN} is the length of the HN bond (1.02 × 10^{-8} cm); τ_c is the correlation time for HN vector.

In the case of rigid particles, where the HN vectors are fixed, τ_c value for a single HN vector is equal to the molecular tumbling, assuming an isotropic mode. The time scale for protein molecular tumbling falls into the range of 10ns. If conformational dynamic for local backbone amides on time scale comparable faster than the molecular tumbling, i.e., ps-ns time scale, ^{15}N R_1 values for the corresponding residues are characterized as higher values than the average one. Accordingly, NOE and R_2 values should be lower than the average one. If motions on μs-ms time scale are present, the modulation of isotropic chemical shifts contributes to the R_{ex} term, resulting in much higher R_2 values. A detail example in charactering local motions of apo and Ca^{2+}-form of S100A5 were shown in Result section 3.3.

2.3 Application of pcs and rdc in studying structural and dynamic properties of protein.

2.3.1 Definition of pcs and rdc

Pcs, which arise in the presence of magnetic susceptibility anisotropy, are given by the following equation:

$$\delta_i^{\mathrm{pcs}} = \frac{1}{12\pi r_i^3}\left[\Delta\chi_{\mathrm{ax}}\left(3\cos^2\vartheta_i - 1\right) + \frac{3}{2}\Delta\chi_{\mathrm{rh}}\sin^2\vartheta_i\cos 2\varphi_i\right] \qquad \textbf{(7)}$$

where $\Delta\chi_{\mathrm{ax}}$ and $\Delta\chi_{\mathrm{rh}}$ are the axial and rhombic anisotropy parameters of the magnetic susceptibility tensor of the metal,

$$\Delta\chi_{\mathrm{ax}} = \chi_{zz} - \frac{\chi_{xx} + \chi_{yy}}{2} \qquad \textbf{(8)}$$

$$\Delta\chi_{\mathrm{rh}} = \chi_{xx} - \chi_{yy}. \qquad \textbf{(9)}$$

r_i is the distance between the atom i and the metal ion, and ϑ_i and φ_i are the spherical polar angles of atom i with respect to the principal axes of the magnetic susceptibility tensor centered on the metal ion. Due to the dipolar nature of this effect, the pseudocontact shifts depend on the distance between the metal ion and the resonating nuclei, independently of the presence or not of chemical bonds. The maximum distance at which the pseudocontact shifts are measurable depends on the magnitude of the magnetic anisotropy and thus on the metal ion. [21,22]

Five components of the χ tensor can be used as parameters, if Equation 7 is re-written in the lab frame. In order to extrapolate this tensor, at least 5 pcs values are needed. After the complete assignment of the ^{1}H-^{15}N-HSQC spectra of the paramagnetic sample and the analogous diamagnetic, the pcs values of N and ^{N}H, for each amino acid, are easily obtained from the difference of the corresponding chemical shifts of the paramagnetic and diamagnetic forms.

Rdc arise from an incomplete averaging of the spatially anisotropic dipolar couplings between two spins. Paramagnetic rdc occur as a result of the magnetic susceptibility anisotropy, which induce partial orientation at high magnetic fields. ^{1}J ^{15}N-^{1}H splittings of the backbone amides are the most frequently used in investigating protein systems. ^{1}J ^{15}N-^{1}H

splittings of the backbone amides experience a residual dipolar coupling contribution which adds to the scalar ^1J value and is given by

$$\Delta\nu_{\mathrm{RDC}}\ (\mathrm{Hz}) = -\frac{1}{4\pi}\frac{{B_0}^2}{15kT}\frac{\gamma_N\gamma_H\hbar}{2\pi r_{NH}^3}\left[\Delta\chi_{ax}(3\cos^2\theta - 1) + \frac{3}{2}\Delta\chi_{rh}\sin^2\theta\cos 2\Omega\right] \quad \textbf{(10)}$$

where θ is the angle between the ^{15}N-^{1}H vector and the *z* axis of the χ tensor, Ω is the angle which describes the position of the projection of the ^{15}N-^{1}H vector on the *xy* plane of the χ tensor, relative to the *x* axis, and $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are defined as in Equation (7). The equation describing this effect is similar to that of pcs[11-13], but in rdc case, the distance r_{NH} is that between the two coupled nuclei and is usually fixed. Therefore, rdc values are not related at all to the position of the coupled nuclei with respect to either the metal ion or the magnetic susceptibility tensor; instead, they depend only on the orientation of the vector connecting the coupled nuclei [10,12, 23].

2.3.2 N60D mutation used in tuning the Ln^{3+} affinity for CaM

The measurements rdc require that the macromolecule of interest is weakly aligned in the magnetic field. For macromolecules containing paramagnetic metal ions, ^{1}J splittings are often measured for both the paramagnetic system and a diamagnetic analogue. Lanthanide ions (Ln^{3+}) are known to be spectroscopic probes for calcium-binding proteins and stand out for their large and varied paramagnetism arising from unpaired electrons in the f orbitals of their trivalent ions. Therefore, Ca^{2+} ions of calcium binding proteins can be substituted by Ln^{3+}, embedding the Ln^{3+} in a rigid and extended molecular framework of defined three-dimensional structure. In this way, pcs and rdc values induced by different Ln^{3+} could be obtained. However, in practice, not all the Ca^{2+} binding protein is able to bind Ln^{3+} tightly and specifically. In case of CaM, in order to overcome this obstacle, a site direct mutation, N60D, was engineered for CaM. The Ln^{3+} can be selectively bound to the second binding site of the N terminal domain of CaM [5,24].

Pcs values of the N-terminal domain of N60DCaM for different lanthanide derivatives were measured in order to determine the magnetic susceptibility anisotropy tensors; pcs and rdc were also measured on the C-terminal domain to provide information on its relative mobility with respect to the domain hosting the paramagnetic center. Pcs and rdc data were

collected for terbium(III) ,thulium(III) and dysprosium(III) calmodulin were supplemented, analogous data were measured for the α-synuclein adduct, the DAPk peptide adduct and the DRP-1 peptide adduct.

2.3.3 MAP program for investigating interdomain motions by using pcs and rdc values

The pcs and rdc values are given by the average of the values corresponding to the experienced conformations. Both pcs and rdc average when the motions occur on time scales faster than ms and are able to incorporate information on motions within a very broad time scale. When the paramagnetic metal resides in one domain, the observation of a smaller range for the rdc values in the other domain reveals the presence of conformational freedom of this domain with respect to the metal-bearing domain.

In this thesis, an approach was developed for determining the maximum allowed probability (MAP) of any conformation in a protein constituted by domains not rigidly connected by using experimental pcs and rdc data. This MAP value describes the largest probability of the protein in a given conformation.

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 Pmax.[25] 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 provide information only on orientation. In this approach, the Pmax values (termed as MAP(R)) were used as the starting point, to which translational information was added by introducing pcs data.

The calculation procedures are as followed. The MAP(R) values relative to all orientations of one domain with respect to the other domain were first obtained using the rdc values. Then, a fit was performed starting from selected orientations R_0 with largest MAP(R_0) values, complemented by other N conformation, with weight (w_i), position (t_i) and orientation (R_i) obtained in order to minimize the target function (TF)

$$TF\ (w_0) = \min_{t_0,(w_i,t_i,R_i)} \sum_j \left| \tilde{\delta}_j - \left(w_0 \delta_j(t_0, R_0) + \sum_{i=1}^{N} w_i \delta_j(t_i, R_i) \right) \right|^2 \qquad \textbf{(11)}$$

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 to N conformations. In order to achieve the absolute minimum for the $TF(w_0)$ (which depends on the number of different metal ions), a high enough number N of conformations should be considered. For the present calculations, relative to three metal ions employed, N was fixed to 9. Such 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 and from rdc 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 $TF(w_0) = \varepsilon$, where ε is the threshold fixed for the error. This was set to a 10% larger value of the absolute minimum of the TF.

Because rdc values are independent of reflections of the axes of the magnetic tensor, the same MAP(R) will be attained for a given orientation as well as for other 3 symmetric orientations, or ghost orientations, which cannot be discriminated. Two or more different metal ions with significantly different magnetic susceptibility anisotropy tensors should be able to eliminate the ghost orientations. However, the problem is that even if two or more paramagnetic ions are considered, the ghosts are not completely removed, 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 with pcs could remove some of the ambiguities, in addition to providing further information on the relative position(s) of the domains.

2.3.4 Refine crystal structure by using pcs and rdc data

Protein solution structures obtained by NMR are limited in precision by the low information content of the experimental restraints and by their small number relative to the

degrees of conformational freedom of the system. Crystal structures are intrinsically more precise. However, crystal structures can suffer from crystal packing forces, so that they may not be accurate models for the structures in solution.

In this thesis, a strategy was developed to improve the accuracy of a protein structure in solution is to take a crystal structure as a starting model and to "correct" it by applying pcs and rdc data. The systems investigated here are also the complex formed between CaM and some of its partners. The pcs and rdc data are induced by Ln^{3+} ions bound to N60D-CaM. The steps of this approach are following:

1) The parameters of the magnetic susceptibility tensor were obtained by fitting pcs data of the metal bearing domain (N-terminal of CaM). The crystal structure was used as the structure model for fitting. The agreement between calculated and observed pcs provides a first indication of the extent of deviation of solution and solid state structures.
2) The ranges of rdc data were used to judge whether the two domains of CaM are fixed or relatively free to move. If the rdc range of the metal bearing domain is similar to that of the other domain, no domain motion is present between the two domains, thus making it possible the structural determination of the system. If the range of the rdc in the other domain is much smaller than that is the metal-bearing one, the two domains are moving with freedom. In this case, MAP program is suitable for describing the probability of the different conformations in space (see section 2.3.3).
3) In case of fixed domains, the crystal structure was refined against pcs and the rdc values from non mobile residues through the routines PARArestraints of Xplor-NIH.[26] The magnetic susceptibility tensor was extrapolated from pcs data of the metal-bearing domain, because pcs are scarcely affected by local mobility and able to provide an accurate estimate of the magnetic susceptibility anisotropy tensor. However, rdc are strongly affected by mobility. The use of rdc for the present purpose should, thus, be only restricted to those coming from groups that do not show evidence of large local motions. To discard the rdc values for the groups affected by mobility, R_1 and R_2 of the involved heteronucleus (^{15}N in this case) were measured.

In the case of CaM complexes with the DAPk and DRP-1 peptides, structural rearrangement were observed in the refined structure with respect to the crystal structure (see result section 3.2).

Reference List

(1) van Tilborg, P. J. A.; Czisch, M.; Mulder, F. A. A.; Folkers, G. E.; Bonvin, A. M. J. J.; Nair, M.; Boelens, R.; Kaptein, R. *Biochemistry* **2000**, *39*, 8747-8757.

(2) Ishima, R.; Torchia, D. A. *Nature Struct.Biol.* **2000**, *7*, 740-743.

(3) Lee, A. L.; Kinnear, S. A.; Wand, A. J. *Nature Structural Biology* **2000**, *7*, 72-77.

(4) Ottiger, M.; Zerbe, O.; Guntert, P.; Wuthrich, K. *Journal of Molecular Biology* **1997**, *272*, 64-81.

(5) Bertini, I.; Donaire, A.; Jiménez, B.; Luchinat, C.; Parigi, G.; Piccioli, M.; Poggi, L. *J.Biomol.NMR* **2001**, *21*, 85-98.

(6) Bertini, I.; Del Bianco, C.; Gelis, I.; Katsaros, N.; Luchinat, C.; Parigi, G.; Peana, M.; Provenzani, A.; Zoroddu, M. A. *Proc.Natl.Acad.Sci.USA* **2004**, *101*, 6841-6846.

(7) Horst, R.; Damberger, F.; Luginbuhl, P.; Guntert, P.; Peng, G.; Nikonova, L.; Leal, W. S.; Wuthrich, K. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 14374-14379.

(8) Cornilescu, G.; Delaglio, F.; Bax, A. *J Biomol NMR* **1999**, *13*, 289-302.

(9) Reif, B.; Hennig, M.; Griesinger, C. *Science* **1997**, *276*, 1230-1233.

(10) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc.Natl.Acad.Sci.USA* **1995**, *92*, 9279-9283.

(11) Tjandra, N.; Grzesiek, S.; Bax, A. *J.Am.Chem.Soc.* **1996**, *118*, 6264-6272.

(12) Bax, A. and Tjandra, N. High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium. J.Biomol.NMR 10, 289-292. 1997.
Ref Type: Journal (Full)

(13) Vold, R. R.; Prosser, P. S. *J.Magn.Reson.Ser.B* **1996**, *113*, 267-271.

(14) Cordier, F.; Grzesiek, S. *J.Am.Chem.Soc.* **1999**, *121*, 1601-1602.

(15) Bertini, I.; Luchinat, C.; Parigi, G. *Solution NMR of paramagnetic molecules*; Elsevier: Amsterdam, 2001.

(16) Liang, B. Y.; Bushweller, J. H.; Tamm, L. K. *Journal of the American Chemical Society* **2006**, *128*, 4389-4397.

(17) Mittermaier, A.; Kay, L. E. *Science* **2006**, *312*, 224-228.

(18) Henzler-Wildman, K. A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M. A.; Petsko, G. A.; Karplus, M.; Hubner, C. G.; Kern, D. *Nature* **2007**, *450*, 838-U13.

(19) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G., III; Skelton, N. J. *Protein NMR Spectroscopy. Principles and practice*; Academic Press: San Diego, 1996.

(20) Abragam, A. *The Principles of Nuclear Magnetism*; Oxford University Press: Oxford, 1961.

(21) Bertini, I.; Luchinat, C.; Parigi, G.; Pierattelli, R. *ChemBioChem* **2005**, *6*, 1536-1549.

(22) Bertini, I.; Luchinat, C.; Parigi, G. *Progr.NMR Spectrosc.* **2002**, *40*, 249-273.

(23) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Nature Struct.Biol.* **1997**, *4*, 292-297.

(24) Bertini, I.; Gelis, I.; Katsaros, N.; Luchinat, C.; Provenzani, A. *Biochemistry* **2003**, *42*, 8011-8021.

(25) Gardner, R. J.; Longinetti, M.; Sgheri, L. *Inv.Probl.* **2005**, *21*, 879-898.

(26) Banci, L.; Bertini, I.; Cavallaro, G.; Giachetti, A.; Luchinat, C.; Parigi, G. *J.Biomol.NMR* **2004**, *28*, 249-261

3
RESULT

During the three years of the PhD course, I mainly focused my work on monitoring the relative orientation of CaM, refining the X-ray structure of the CaM-peptide complex by using paramagnetic constraints and the structure determination and dynamic studies on the S100A5 protein. All these work have been performed by NMR spectroscopy.

As discussed in the introduction section, CaM is a peculiar protein, which has two domains essentially independent in solution. X-ray investigations provide detailed pictures of frozen conformations in the solid state among the many possible in solution, whereas NMR has provided a wealth of dynamics information. I designed a novel program, called MAP, with Fortran 77 language a program, which was applied to provide maximum allowed probabilities (MAPs) of conformations in protein domains not rigidly connected. It is based on the characterization of the conformational space in terms of a maximum probability value 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 CaM, N60D mutant, as well as to its adduct with α-synuclein. The pcs and rdc values were obtained by using the data measured for Tb^{3+}, Tm^{3+}, and Dy^{3+}, when bound to CaM. From the MAP analysis, it shows that free CaM experience a large number of conformations. In the adduct with α-synuclein, CaM still adopts a large ensemble of conformations, however, the conformations with the largest MAP values are in a region of space close to ‘closed conformation’, which is similar to those observed for CaM interacting with peptides with high affinity. In this work, my contribution is designing program. The NMR data were obtained from published work. Details were reported in result section 3.1

For the rigid system, we developed a strategy in order to improve the accuracy of a protein structure in solution by taking a relatively good crystal structure as a starting model and “correct” it by pcs and rdc values. To judge whether domain motions are present in the multidomian protein, we used the spread of the rdc values. In case of relatively rigid protein, the spread of the rdc values is the same for the different domains, when the paramagnetic center is localized on one of them. On the contrary, the domains are relatively free if the spread of rdc values were different. In case of free CaM, the domain motion is present, while the two domains are rigid when it DAPk and DRP-1 bound. Thus it allows the structural

determination of CaM bound with these peptides. Since the crystal structures of CaM complex with DAPk and DRP-1 peptide have already been available, I refined these structures by using pcs and rdc values. From the refinement it showed that the solution structures are similar, but not identical, to the crystal structures that can be ascribed to a structural rearrangement from solid state to solution. In this work, my contribution is running all the NMR experiments and data analysis. Detailed works were shown in result section 3.2.

In the work of S100A5, I contributed the solution structures determination of both the apo and the Ca^{2+} bound forms of S100A5. Relaxation parameters for both forms were also measured accordingly. Both forms are homodimers. The structural differences induced by Ca^{2+} binding mainly occurs at helix III, helix IV, the hinge loop and the last C-terminal residues, similarly to what found for other S100 proteins. The dynamic properties for S100A5 and comparison with other S100 proteins were also discussed in details. In this work, my contribution is the NMR part and structural determination. Detailed description was shown in result section 3.3.

3.1

Paramagnetism-based NMR restraints provide a maximum probability ranking of the different conformations of partially independent protein domains

Ivano Bertini[1,2], Yogesh K. Gupta[1], Claudio Luchinat[1,3], Giacomo Parigi[1,3], Massimiliano Peana[1], Luca Sgheri[4], Jing Yuan[1]

[1]Magnetic Resonance Center (CERM), University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy. [2]Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy. [3]Department of Agricultural Biotechnology, University of Florence, via Maragliano 75-77, 50144 Florence, Italy. [4]Istituto per le Applicazioni del Calcolo—Sezione di Firenze, Polo Scientifico—CNR, Via Madonna del Piano 10, 50019 Sesto Fiorentino, Italy

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.

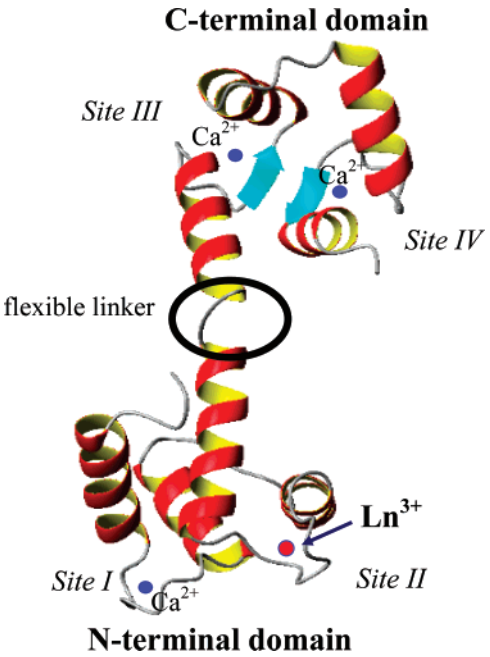


Figure 1. CaM can bind four calcium ions, two in the N-terminal domain and two in the C-terminal domain. The N60D mutant binds lanthanides selectively at the second binding site of the N-terminal domain. The two domains are shown as observed in the so-called "extended" conformation of CaM (PDB 1CLL).

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.

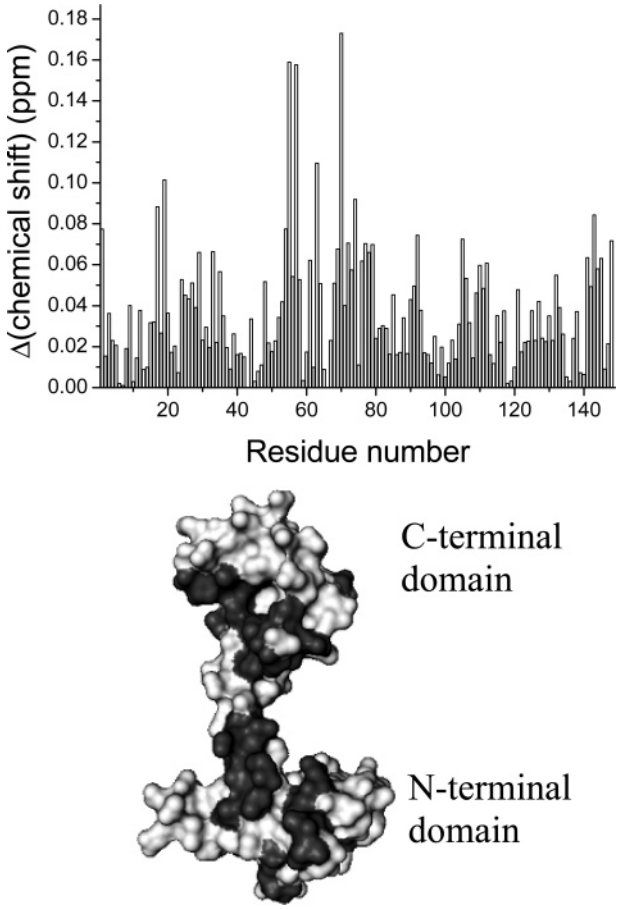


Figure 2. Plot of the change in chemical shift observed for the CaM H^N and N backbone atoms in the 1:1 ^{15}N AS−CaM complex. Shifts are reported as a weighted average of the amide proton and amide nitrogen shifts using the formula $\Delta\delta = [(\Delta\delta_H)^2 + (\Delta\delta_N \times (\{\gamma_N\}/\{\gamma_H\}))^2]^{0.5}$. 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.

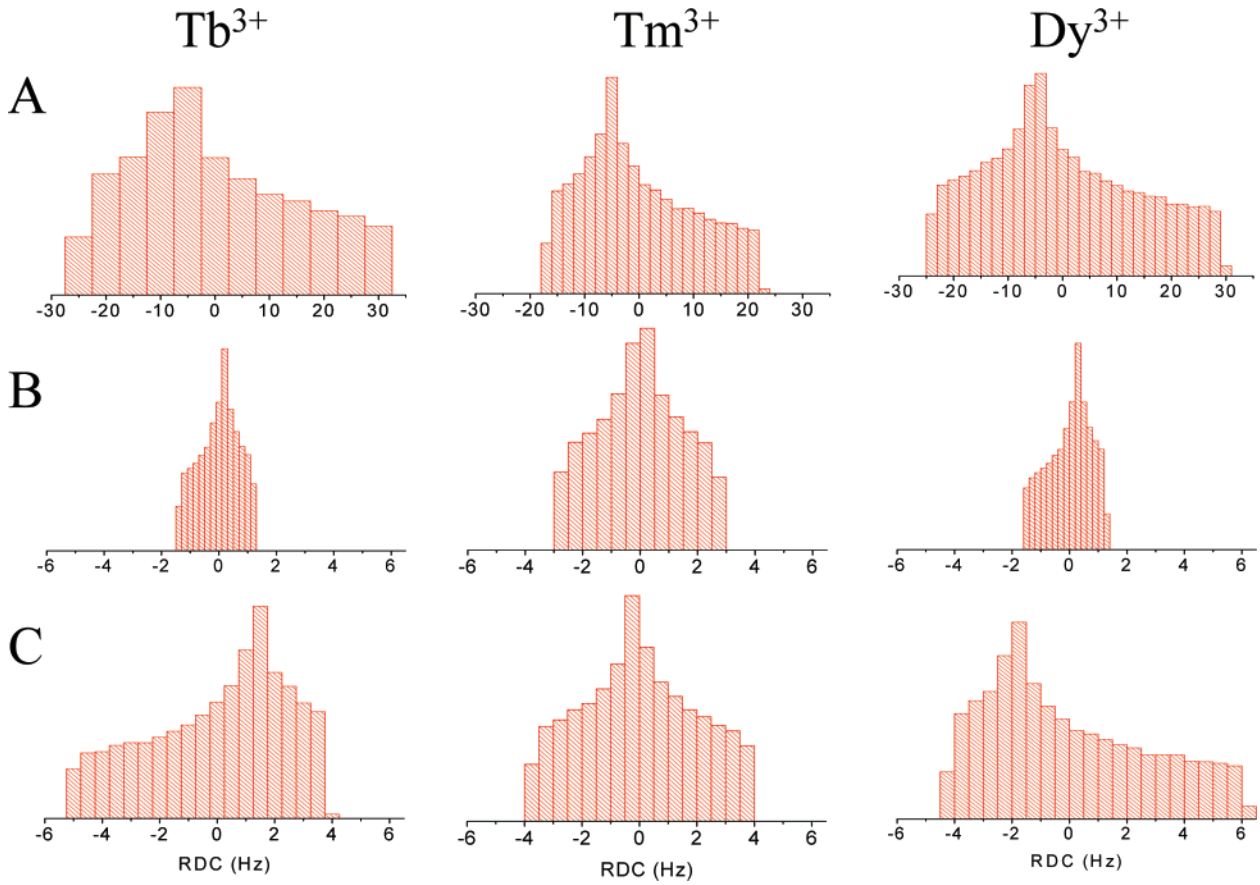


Figure 3. Observed spreading of rdc values in the C terminal domain of $(CaLn)_N(Ca_2)_C$CaM (Ln = Tb, Tm, or Dy) in the free form (B) or complexed with AS (C) as compared with the spreading predicted in the absence of conformational freedom (A).

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.

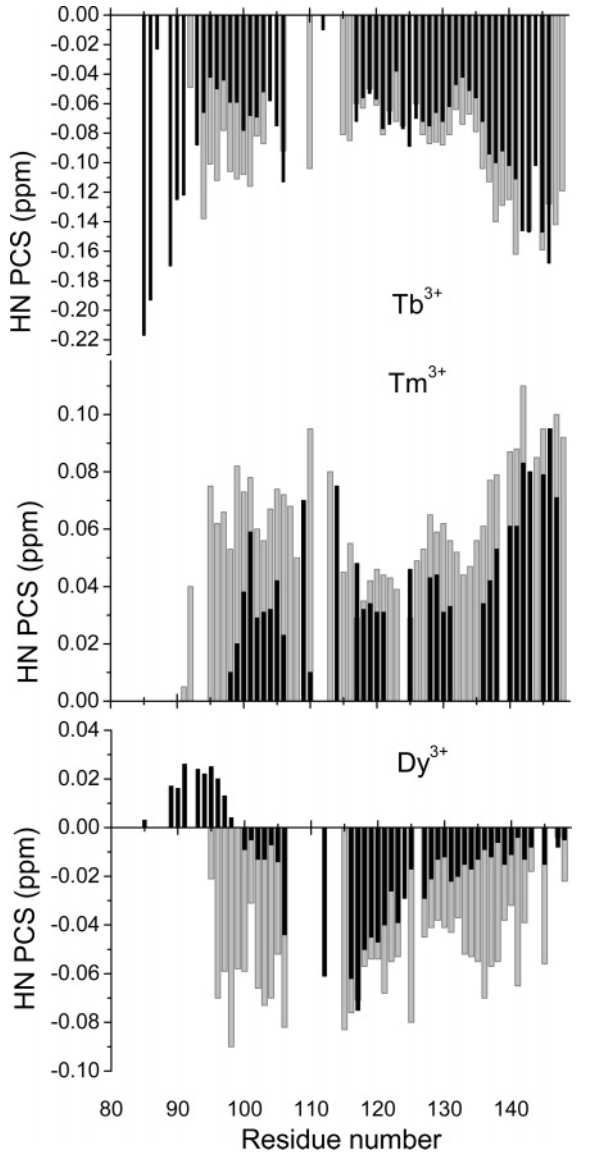


Figure 4. Observed C-terminal HN pcs values for the terbium(III), thulium(III), and dysprosium(III) CaM derivatives in the free form (black) and in the presence of AS (gray).

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

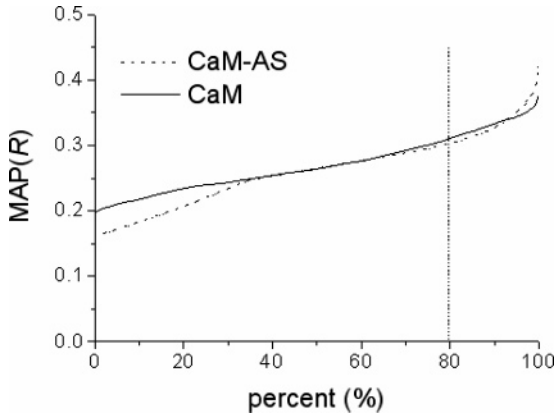


Figure 5. MAP(R) values calculated from rdc data for CaM and CaM−AS. A point (x,y) on the graph means that a fraction x of all orientations have a value of MAP(R) $\leq y$.

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

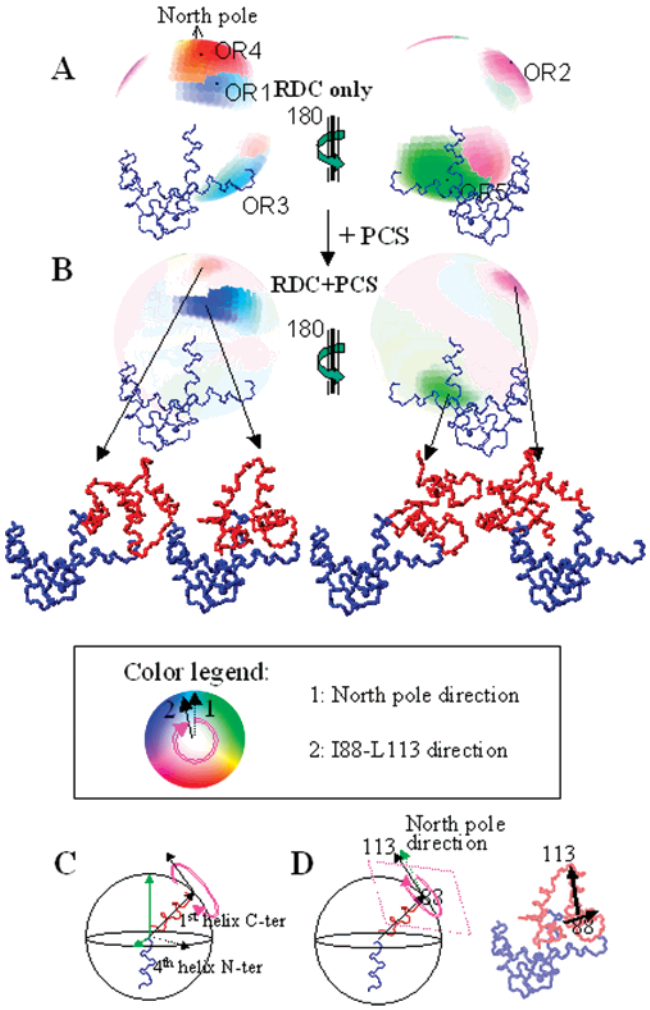


Figure 6. (A) MAP(*R*) values calculated from rdc data for all relative orientations of the C-terminal domain with respect to the N-terminal domain (in blue) of free CaM. The first C-terminal residue and the last N-terminal residue outside the mobile 78−81 hinge region are placed in the center of the 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

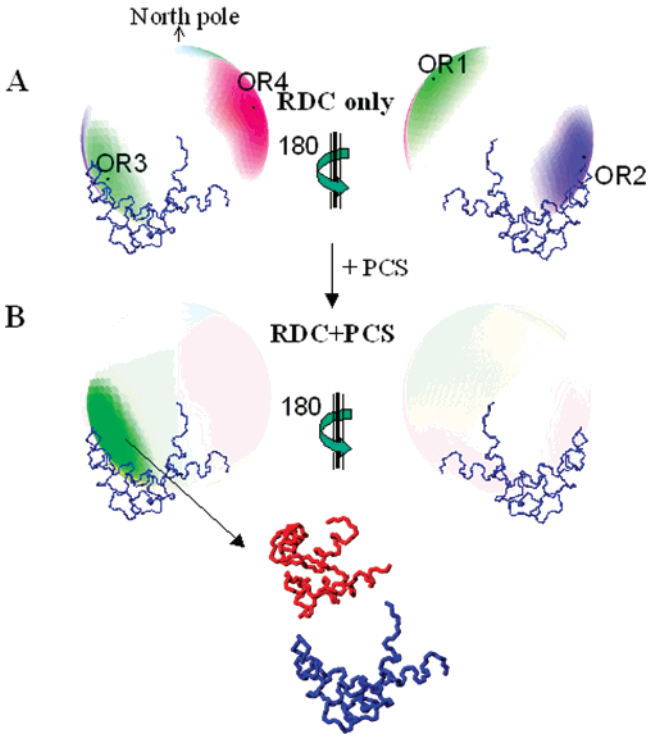


Figure 7. (A) MAP(*R*) values calculated from rdc data for all relative orientations of the C-terminal domain with respect to the N-terminal domain (in blue) of CaM in the presence of AS. (B) Conformations with the largest MAP values in agreement with both rdc and pcs data. Details same as those for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

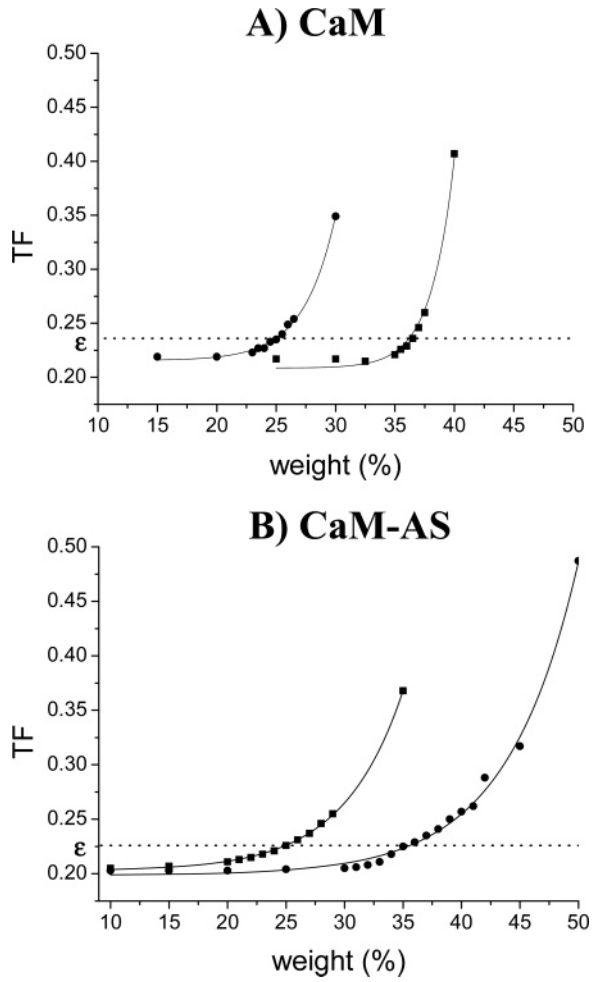


Figure 8. Target function $TF(w_0)$ for two different orientations in bad or good agreement with the experimental data for the CaM (A) and CaM−AS (B) cases. The TF function has a roughly exponential behavior, as shown by the fits. The maximum weight corresponding to a *TF* value equal to ϵ (shown as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

3.2

Accurate solution structures of proteins from X-ray data and a minimal set of NMR data: calmodulin-peptide complexes as examples

Ivano Bertini[1,2*], Petri Kursula[3,4], Claudio Luchinat[1,5], Giacomo Parigi[1,5], Juha Vahokoski[3], Matthias Wilmanns[3], Jing Yuan[1]

[1]Magnetic Resonance Center (CERM), University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy.

*e-mail: ivanobertini@cerm.unifi.it

[2]Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy.

[3]EMBL-Hamburg c/o DESY, Hamburg, Germany

[4]Department of Biochemistry, University of Oulu, Oulu, Finland

[5] Department of Agricultural Biotechnology, University of Florence, via Maragliano 75-77, 50144 Florence, Italy

Abstract

A strategy for the accurate determination of protein solution structures starting from X-ray data and a minimal set of NMR data is proposed, and successfully applied to two complexes of calmodulin (CaM) with target peptides not previously described. Its implementation in the present case is based on the use of lanthanide ions as substitutes for calcium in one of the four calmodulin binding sites, and the collection of pseudocontact shifts (pcs) and residual dipolar coupling (rdc) restraints induced by the paramagnetic metals. Starting from the crystal structures, new structural models are calculated that are in excellent agreement with the paramagnetic restraints and differ modestly but significantly from the starting crystal structures. In particular, in both complexes, a change in orientation of the first helix of the N-terminal CaM domain and of the whole C-terminal domain is observed. The simultaneous use of paramagnetic pcs and rdc restraints has the following crucial advantages: i) it allows one to assess the possible presence of interdomain conformational freedom – which cannot be detected if the rdc are derived from external orienting media; ii) in the absence of significant conformational freedom, the global orientation tensor can be independently and precisely determined from pcs, which are less sensitive than rdc to the presence of local structural disorder or mobility and iii) the *relative* rearrangement of a domain or a secondary structure element with respect to the metal-bearing domain can be accurately detected.

Introduction

Protein solution structures obtained by NMR are limited in precision by the low information content of the experimental restraints and by their small number relative to the degrees of conformational freedom of the system. Furthermore, local and global atomic movements undermine the quantitative reliability of all restraints, as none of them is a simple function of the movement. As a result, a *family* of solution structures is usually provided that are all consistent with the experimental restraints, and whose spread provides but a qualitative measure of the precision of the structure itself.[1] The RMSD among the members of the family is relatively high (typically 0.5-1 Å for backbone atoms and higher for side chain atoms).

Crystal structures are intrinsically more precise: while the average precision of conventional NMR structures corresponds to a resolution of approximately 3 Å,[2;3] the precision of an "atomic resolution" crystal structure corresponds to about 1 Å resolution. Yet, solid state structures can suffer from crystal packing forces - and sometimes from being recorded at liquid nitrogen temperature - so that, independently of the precision, they may not be an *accurate* model for the structure in solution. In other words, the "true" solid state structure may be different from the "true" solution structure. In such cases, for proteins that perform their *in vivo* function in solution, the latter and not the former must be the target of any structural study. The current situation is illustrated in Figure 1A. This has been extensively discussed in the literature, especially for multiple-domain proteins.[4-6]

A possible strategy to improve the accuracy of a protein structure in solution is to take a relatively good crystal structure as a starting model and to "correct" it by applying well selected, sensitive NMR restraints. Such correction could generate a structure that is closer to the "true" solution structure, i.e. more accurate than the starting X-ray structure model, and at the same time more precise than the solution structure obtained with conventional NMR methods (Figure 1B). An early example of this strategy is provided by the use of pseudocontact shifts (pcs) from a paramagnetic ion in a metalloprotein.[7] More recently, residual dipolar couplings (rdc) originating from external orienting media have been proposed and highly successfully used.[4;5]

We propose here a strategy based on the occurrence of a paramagnetic metal ion binding site - either natural or artificial - and the *combined* use of paramagnetism-based pcs and rdc. The latter arise from partial self-orientation due to paramagnetic susceptibility anisotropy, $\Delta\chi$, so no external orienting media are needed. Both pcs and rdc are long range restraints and, therefore, optimally suited to detect global structural features, especially relative orientations of secondary structure elements or entire domains.[4;8-11] These are precisely the features that, when altered by crystal packing forces, may make crystal structures inaccurate models of

solution structures. Rdc are also strongly affected by mobility.[12-14] The use of rdc for the present purpose should, thus, be only restricted to those coming from groups that do not show evidence of large local motions. To discard the rdc values for the groups affected by mobility, R_1 and R_2 of the involved heteronucleus (^{15}N in this case) can be measured.

If a structural model is available, as is the case for crystal structures, the pcs provide an accurate estimate of the magnetic susceptibility anisotropy tensor, and the rdc can be used to refine the structure through the latter tensor. Pcs and rdc are in fact differently sensitive to local and global motions. Pcs are mostly sensitive to large global movements, and are scarcely affected by local mobility.[9;10;15] Rdc, on the other hand, are sensitive to even small domain reorientations. The agreement between calculated and observed pcs provides a first indication of the extent of deviation of solution and solid state structures. A new structural model using pcs and rdc can then be calculated, and both tensor and structure can be refined iteratively.

It should be stressed that the use of pcs and rdc originating from a protein-bound paramagnetic metal is intrinsically different from using pcs restraints alone on one hand[7] and rdc restraints induced by external media on the other.[4;5] In fact, the independent availability of an accurate estimate of the orientation tensor from pcs permits a more quantitative use of rdc themselves: when rdc are initially used to assess the quality of a structural model and the orienting tensor is known (and fixed), it should be easier to discriminate the contribution to the deviations originated by small local rearrangements from those originated by global reorientations of domains or subdomains. This is particularly true for two-domain (or multi-domain) proteins. As the metal resides in one domain, the observation of a smaller range for the rdc values in the other domain immediately reveals the presence of conformational freedom of this domain with respect to the metal-bearing domain.[9;10] On the other hand, if the two sets of rdc have similar magnitudes, even modest differences in the orientation of the rdc-derived tensor in the neighboring domain with respect to that in the metal-bearing domain should reveal a change in the relative orientation of the two domains with respect to the starting crystal structural model. This is at striking variance from a situation where the orientation tensors originate from external orienting media. In this case, the magnitudes of the two orienting tensors are unrelated to the conformation freedom. A limit case is provided by calmodulin (CaM), a protein with two domains (each constituted by a pair of helix-calcium binding loop-helix EF-hand motifs, and connected by a linker) which binds target peptides by wrapping the two domains around the target, but where the two domains are free to move with respect to one another in the peptide-free form.[4;16;17] In the case of free CaM, while in the presence of a paramagnetic metal bound to one domain the rdc-derived tensor from the other domain is more than one order of magnitude smaller,[9] in the

presence of external orienting media the rdc-derived tensors are similar for the two domains (Figure 2).[4] Obviously, if one makes a wrong assumption about the absence of conformational freedom in free CaM, the two tensors could be made to essentially coincide by rotating one domain with respect to the other, providing a structure devoided of physical meaning. Paramagnetic NMR has provided a wealth of dynamics information on free CaM,[9] recently providing a maximum allowed probability (MAP) for each reciprocal orientation of the two domains.[10]

The obvious test of the present strategy is therefore the solution structural characterization of CaM in its complex with two peptides representing the interaction sequence of two protein partners, the death-associated protein kinase (DAPk) and the DAPk-related protein 1 (DRP-1).[18] Binding protein targets or their relevant peptides tends to restrict the freedom of one domain with respect to the other, possibly until freezing of a single conformation, depending on the strength of the interaction.[19] X-ray investigations provide detailed pictures of frozen conformations in the solid state among the many possible in solution, which are, however, somewhat different from one peptide to another.[19] It is, thus, possible that i) some conformational freedom of the two domains is maintained in solution, as it has been observed in at least three cases[10;20;21] and is not revealed by the solid state structures, and ii) even if the structure of the complex is also immobilized in solution, the relative orientation of the two domains in the solid state is determined by relatively weak forces, possibly of the order of the crystal packing forces, so that differences between solid state and solution structures may be expected. Even the relative orientation of the helices constituting each EF-hand domain can vary[4] and be another point of possible discrepancy between solid state and solution structures.

We therefore solved the high-resolution crystal structures of these two novel CaM-peptide complexes; then we applied our strategy to substitute the calcium ion in the second EF-hand of the N-terminal domain with paramagnetic lanthanide ions using the N60D mutant (three lanthanide ions were used in this case).[22] As discussed above, if the two domains are fixed in a rigid structure, the range of the observed NH residual dipolar coupling values (rdc) should be the same for the two domains; otherwise it should be smaller for the C-terminal domain, as the latter would experience averaging of the contributions from the different relative orientations of the N-terminal domain or from complete rotational freedom.[9] In case each protein domain can be considered rigid, NH rdc referred to one lanthanide are in principle enough to detect changes in their relative orientation; at least two metals are necessary if a local domain minimization is performed while minimizing their relative orientation. Three lanthanides were here used to strengthen the results and avoid possible non

global minima. This issue is further discussed in the Results and Discussion section.

In the present complexes, the two domains appear essentially rigidly fixed. Therefore, this is a case where the present strategy can be fully deployed. A careful analysis of the rdc indicates differences in the solution structure relative to the crystal structure, namely modest but significant changes in the orientation of helix 1 in the first EF-hand of the N-terminal domain, and of the whole C-terminal domain, with respect to the metal-bearing second EF-hand of the N-terminal domain. This new approach thus provides solution structures that are of the same quality of the starting solid state structures and more accurate of the latter as models for the true solution structures.

Materials and Methods

Sample preparation, crystallization, and structure solution. Human CaM was expressed in *E. coli* and purified by Ca-dependent hydrophobic interaction chromatography on phenyl sepharose. The peptides RKKWKQSVRLISLCQRLSR and RRRWKLSFSIVSLCNHLTR, representing the amino acid sequences of the CaM-binding domain of DAPk and DRP-1 (residues 302-320), were purchased from Mimotopes. Crystallisation was carried out at +20°C by vapour diffusion, in sitting drops containing 1 µl of protein-peptide mixture (0.5 mM CaM, 1.5 mM peptide in 20 mM $CaCl_2$, 50 mM HEPES, pH 7.5) and 1 µl of well solution. The optimal well solution contained 30 % PEG 1500, 10 mM DTT, and 0.1 M sodium acetate (pH 4.8). Data were collected at 100 K on beamline X13 at EMBL-Hamburg/DESY. The diffraction data were processed using XDS[23] and XDSi[24]. Initial phasing was carried out by molecular replacement in MOLREP,[25] using the N- and C-terminal domains of CaM separately. Refinement was performed using REFMAC5[26] with TLS parameters.[27] Model building and analysis were done in O.[28] Water molecules were added both manually and with Arp/Warp.[29] Residues 2-148 of CaM and all residues of the peptide were built into the model. The processing and refinement statistics are shown in tables S1-S2 in the Supporting Information. The coordinates and structure factors were deposited at the Protein Data Bank with accession codes 1WRZ (DRP-1) and 1YR5 (DAPk).

^{15}N and ^{13}C labeled N60D CaM was purchased from ProtEra. NMR samples of Ca_4CaM and $LnCa_3$CaM (Ln=Tb, Tm, Yb and Dy) were prepared as previously reported[9;10] (HEPES 30 mM, KCl 200 mM, TCEP 3 mM pH=7.4). CaM concentration was 0.5 mM, peptides were in slight excess.

NMR Measurements. ^{1}H ^{15}N HSQC experiments were performed at 700 MHz. Pcs data were obtained as the ^{1}H and ^{15}N chemical shift difference between the paramagnetic form and the

diamagnetic form. Rdc data were obtained from IPAP experiments[30] at 700 MHz. HNCO, HNCA, HN(CO)CA and CBCA(CO)NH spectra for backbone assignment were acquired on a Bruker 500 MHz spectrometer equipped with a cryoprobe. ^{15}N relaxation rates (R_1 and R_2) were measured at 70.94 MHz ^{15}N base frequency using standard pulse schemes[17;31] to collect ten points with delays of 2.5, 75, 125, 275, 400, 500, 600, 850, 1500, 2000 ms for R_1 and nine points with delays of 16.96, 33.92, 50.88, 67.84, 84.80, 118.72, 152.64, 186.56, 237.44 ms for R_2. Relaxation rates were determined by fitting the crosspeak heights, obtained through the standard routine of the Sparky program.[32] All experiments were performed at 298 K.

Paramagnetism-based restraints. The electron-nucleus dipolar coupling does not average to zero upon rotation in the presence of anisotropy in the paramagnetic susceptibility tensor. A contribution to the hyperfine shift, which is called *pseudocontact shift* (pcs) thus arises, which is described by Eq. (1)[33]

$$pcs = \frac{1}{12\pi r^3}\left[\Delta\chi_{ax}\left(3\cos^2\theta - 1\right) + \frac{3}{2}\Delta\chi_{rh}\sin^2\theta\,\cos 2\varphi\right] \qquad (1)$$

where r is the distance between observed nuclei and metal ion, $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are the axial and rhombic anisotropy parameters of the susceptibility tensor of the metal, and θ and φ identify the polar coordinates of the nucleus in the frame of the paramagnetic susceptibility tensor. Therefore, pcs values depend only on the position of the nuclei with respect to both the metal ion and the paramagnetic susceptibility tensor, besides the value of the anisotropies of the latter.

The ^1J splittings of coupled nuclei can experience dipolar contributions, due to partial self-orientation of the investigated system in the magnetic field. Such contribution is called residual dipolar coupling (rdc) and is provided by Eq. (2)[33;34]

$$rdc\ (\mathrm{Hz}) = -\frac{1}{4\pi}\frac{B_0^{\,2}}{15kT}\frac{\gamma_N\gamma_H\hbar}{2\pi r_{HN}^3}\left[\Delta\chi_{ax}(3\cos^2\theta - 1) + \frac{3}{2}\Delta\chi_{rh}\sin^2\theta\cos 2\phi\right] \qquad (2)$$

where r_{HN} is the distance between the two coupled nuclei N and ^{N}H and the polar angles θ and ϕ are those defining the orientation of the vector connecting the coupled nuclei in the frame of the magnetic susceptibility tensor. Other symbols have the usual meaning. Therefore, rdc values are not related at all to the position of the coupled nuclei with respect to both the metal ion and the magnetic susceptibility tensor, but they depend only on the orientation of the vector connecting the coupled nuclei in the reference frame of the magnetic susceptibility tensor axes.[35-38] As can be easily seen from Eq. (2), more than one θ and ϕ pair can provide the same rdc value; rdc measurements referring to at least two metals are thus necessary.

Pcs and rdc may act as reporters of structural information because they depend on the position of the observed nuclei (for pcs) or on the orientation of the vector connecting coupled nuclei (for rdc) with respect to the paramagnetic susceptibility anisotropy tensor (see Eqs. 1 and 2).[39] Pcs and rdc may also act as reporters of protein mobility because the magnetic susceptibility anisotropy parameters $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ values obtained for one domain with respect to those of another domain provide information on the relative motion between the two domains.[9;10] Analogously, deviations from the expected pcs/rdc values relative to individual residues may provide information on their motional averaging.

Refining the structure in solution. The protein structures in solution were refined using the crystal structures as starting models and correcting them by applying the NMR restraints.[4;5;40] The pcs and the rdc values from non mobile residues were provided as restraints through the routines PARArestraints for Xplor-NIH.[41] Initially, a rigid minimization was performed, using the pcs restraints, to determine the orientation of the magnetic anisotropy susceptibility tensors of each metal placed at the second binding site of the N-terminal domain. The crystal structure was then subjected to a simulated annealing at 100 K, being restrained at the backbone torsion ϕ and ψ angles extracted from the structure itself, in order to be minimized with respect to the employed library (topology and parameter files were topallhdg5.3 and parallhdg5.3, respectively). This determined a slight rearrangement in the protein structure, the amplitude of which was around 0.4 Å of backbone RMSD. As a second step, an internal dynamics at 200 K and a minimization were performed, with the force constant of the starting backbone torsion ϕ and ψ angle restraints ramped down, and with the addition of the dihedral angle restraints calculated with TALOS with force constant of 1256 kJ mol^{-1} rad^{-2}, of the pcs restraints with force constant of 41.87 kJ mol^{-1} ppm^{-2}, and of the rdc restraints with force constant of 0.837 and 3.35 kJ mol^{-1} Hz^{-2} for Tb^{3+}, Tm^{3+} and Dy^{3+} and for Yb^{3+}, respectively. The force constant for Yb rdc is larger than for the other metals because of the smaller values of Yb rdc. Pcs were thus used in the first step to determine the anisotropy tensor, and in the second step to refine the protein structure together with the rdc. In summary, the protocol consists in a refinement of the crystal structure performed at low temperature in order to have the smallest changes required for the best agreement of all the experimental paramagnetism-based restraints.

Results and discussion

The crystal structures

The crystal structures of Ca^{2+}-bound CaM complexed with the DAPk and the DRP-1

peptides were determined (see the Materials and Methods section) at 1.7 and 2.0 Å resolution (PDB 1YR5 and 1WRZ), respectively. The N-terminal and the C-terminal CaM domains wrap around the bound peptide, which has an α-helical structure, according to the canonical closed state. The peptides bind in an antiparallel orientation, i.e. with the N-terminal and the C-terminal CaM domains interacting mainly with the C-terminal and N-terminal halves of the peptide, respectively. The same orientation was observed for peptides derived from MLCK[42;43] and from CaM-dependent protein kinase II.[44]

The complex is stabilized by several hydrophobic interactions. Tryptophan 305 and leucine 318 mainly act to anchor the peptide to the hydrophobic patches of CaM. In addition to the hydrophobic interactions there are a number of possible electrostatic interactions, such as those between peptide arginines and lysines and glutamate CaM residues in helix I and VII.

The mechanism of activation by CaM is common for all its protein-derived peptide targets. However, four different recognition modes have been identified, termed 1-10, 1-14, 1-16 or 1-17 motifs based on the position of the two key anchoring hydrophobic residues in the target peptide.[19;45] A fifth interaction motif has been described for the gating domain of the small conductance Ca^{2+}-activated K^{+} channels.[46] The recognition mode for DAPk and DRP-1 corresponds to the 1-14 motif. The same motif has been identified for the skeletal muscle and smooth muscle MLCK peptides[42;43] (PDB 1CDL), the endothelial nitric oxide synthase peptide[47] (1NIW), and a peptide derived from the olfactory CNG channel[48] (1SY9). A comparison among the CaM structures in these complexes with respect to the structures determined for the complex with DAPk or DRP-1 peptides shows that all structures are in an overall agreement (see Figure 3). However, although the three 1CDL, 1NIW and 1SY9 structures show the central CaM linker differently unwrapped, all of them provide about the same relative orientation between the last helix of the N-terminal domain and the first helix of the C-terminal domain. Figure 3, on the contrary, clearly shows that in the complex with the DAPk or DRP-1 peptides the first helix of the C-terminal domain is oriented differently from the other cases. Therefore, all residues in the interdomain linker, as well as those of the C-terminal domain up to residue 87, have a large RMSD with the 1CDL, 1NIW and 1SY9 structures. Figure 3 also shows that the first helix of the N-terminal domain can be differently oriented depending on the bound peptide. It has already been noted that this depends on the specific interactions between the binding peptide and the first helix residues.[47]

Of relevance to the present study are the intermolecular interactions that connect each CaM peptide complex with the neighboring molecules in the crystal. Such interactions, lacking in solution, deserve a systematic analysis, as they may be the cause of any possible structural difference between the crystal structure and the solution structure of the complexes.

Table 1 lists all H-bond, salt bridge and Van der Waals contacts between each complex and its neighbours in the crystalline state. It can be noted that a number of hydrogen bonds are present between the N-terminal and neighbour molecules 3, 5 and 7, those with molecule 3 involving helix 1, those with molecule 5 involving helix 3 and those with molecule 7 involving helices 1 and 4 of the N-terminal domain. These interactions are essentially the same in both complexes. In addition, a salt bridge between glutamates 7 and 14 and arginine 2 in DAPk, and between glutamates 7 and 14 and arginine 3 as well as between glutamate 11 and lysine 6 in DRP-1 establish contact between the N-terminal domain and the peptide, helping the maintenance of the closed conformation of the complex. A summary of the relevant interactions is schematically shown in Figure 4.

Choice and collection of paramagnetic restraints

A few considerations are needed on the choice of the most appropriate restraints for this type of study. It is always desirable to have a variety of rdc datasets for either structure determination or refinement.[4;33;49] In the present case, it is also desirable to have more than one dataset of pcs. The latter can only be obtained by the use of more than one lanthanide. For each lanthanide, several rdc datasets can be acquired, providing information on different internuclear vectors. In principle, a large number of different rdc datasets can be obtained. On the other hand, for the present method to have practical appeal, a too extensive collection of experimental data should be avoided.

From the above considerations, we resorted to using *three* different lanthanides and to collecting pcs data for the three of them and rdc data only for the backbone N-H vectors. This choice has several advantages: i) preparation of different lanthanide derivatives is a trivial task given the chemical similarity among lanthanides, and this provides independent pcs datasets (the assignment of the corresponding spectra is also straightforward)[39;50], ii) measuring only N-H rdc has the advantage than singly labelled protein samples can be used, iii) the ratio between paramagnetism-based rdc values and the line-broadening on the heteronuclear dimension is most favourable for N-H vectors.[39;51] In addition, for proteins with alpha-helical secondary structure, such as calmodulin, the N-H vectors, among all vectors for which rdc can be measured, are the most suitable to detect global helical moviments, being almost parallel to the helix axis.

If we are dealing with global movements of rigid domains, even a single rdc dataset could be enough to describe the movement, provided that the other three symmetry-related solutions can be discarded because strongly inconsistent with the starting model. However, local vector rearrangements with respect to the starting crystal structure are also to be

expected. If the degrees of freedom for each N-H vector were totally independent from the other N-H vectors, as it would be necessary to assume in the absence of a starting structural model, at least three independent datasets would be needed. In the present case, it is reasonable to assume that the degrees of freedom of each N-H are only partially independent from one another, and two independent rdc datasets may be sufficient. We have decided to use three lanthanides, and therefore three independent rdc datasets, but we have also checked that even two of the three datasets are often enough to approximate the correct final structure, as it will be shown below.

NMR experiments were performed on the N60D variant of CaM, because in this variant the second binding site of its N-terminal domain can selectively bind a paramagnetic lanthanide ion.[9] It has already been shown that Ln-substitution and the N60D mutation do not affect the protein structure besides the metal coordination sphere.[9;10;22] As far as long range electrostatic interactions are concerned, the substitution of the tripositive Ln^{3+} ion for the Ca^{2+} ion is compensated by the additional negative charge introduced next to the metal by the N60D mutation. Pcs of ^{N}H, N nuclei and rdc of the ^{N}H-N pairs for three paramagnetic $LnCa_3CaM$ forms when bound to the DAPk (Ln = Tb, Tm, Yb) or to the DRP-1 (Ln = Tb, Tm, Dy) peptide were measured. All the following analysis is essentially based on three ^{N}H pcs and three ^{N}H-N pairs rdc datasets for each of the two complexes. Additional rdc data for a fourth Ln derivative of the DAPk complex (Dy), pcs data for N, C^{α}, C^{β}, C' for both complexes, and two sets (Ln = Tm, Yb) of C^{α}–H^{α} pairs rdc for the DAPk complex were also collected, and used as detailed below.

Initial determination of the magnetic anisotropy tensors from pcs and consistency of the rdc data

According to the proposed strategy, the magnitude and orientation of the Ln-centered magnetic susceptibility tensors is initially calculated by fitting the amide H pcs values of the N-terminal domain (the domain bearing the paramagnetic metal) to the crystal structure (Figure 5A,B and Table 2). Then a first check for consistency of the rdc data with the derived tensors is made. If the rdc data are reproduced well, it can be immediately concluded that the solution structure does not differ appreciably from the crystal structure. Pcs are indeed quite robust to provide the magnetic susceptibility anisotropy tensors, because they are not very sensitive to small local conformational changes, and the error in the measurement of amide proton pcs is indeed quite small (0.1 ppm).[52] To be noted that, although the magnetic anisotropies for Yb are clearly smaller than those of the other three lanthanide ions, the range of the pcs values observed for Yb is similar to that of the other ions because pcs values for

atoms closer to the lanthanide can be measured.

The rdc measured for the C-terminal domain for each of the three metal derivatives and each of the two peptide complexes span ranges of values similar to those measured for the N-terminal domain (Figure S1), differently from the free CaM case, when the rdc measured in the C-terminal domain were much smaller than those of the N-terminal domain. This already suggests that in these complexes the two CaM domains are essentially maintaining a fixed conformation with respect to one another; it should be noted again that this conclusion is straightforwardly drawn because rdc are originated by a paramagnetic ion bound to one domain, but would not have been obtained if external orienting media were used. Differently from pcs, rdc are very sensitive both to small local conformational changes and to mobility, because they depend on the orientation of the N-^{N}H vector with respect to the axes of the magnetic susceptibility anisotropy tensor. The rdc were back calculated by using the tensors obtained from the fit of the pcs data as orienting tensors, and found to disagree with the experimental values for a relatively large number of residues by much more than the experimental error (Figure 5C,D). As a consequence, these residues should be either subject to sizable mobility or experiencing a somewhat different conformation in solution with respect to the solid state, or both. The corresponding Q_{free} values are 0.86, 0.68 and 0.67 for Tb, Tm and Yb rdc, respectively, for the complex with the DAPk peptide, and 0.71, 0.47 and 0.74 for Tb, Tm and Dy rdc, respectively, for the complex with the DRP1 peptide.

Discarding mobile residues

At this point R_1 and R_2 measurements[31] were performed to discard the NHs experiencing mobility in the picosecond to nanosecond and/or microsecond to millisecond time scales (Fig. 6 and Fig. 7). Large mobility is observed for the residues in the linker between the N-terminal and C-terminal domain, and for residues 42, 57, 113-116, 130, 137 and 138 (all in protein loops), consistently with previous measurements performed for CaM complexed with other peptides (see Supporting Information).[17;53-55] A R_2 value or R_2/R_1 ratio larger than that calculated with HYDRONMR[56] is in particular observed for some residues, mainly residues of the first helix of the C-terminal domain and/or interacting with the bound peptide (residues 24, 39, 92, 93, 122, 127, 141, 144 and 145 in the DAPk case and residues 16, 64, 87, 93 and 96 in the DRP-1 case): these residues are thus expected to experience motions in the microsecond to millisecond time scale.[31;57] As already noted for CaM bound to the CaMKI peptide, these data account for significant contributions from chemical exchange phenomena.[53] Further considerations on the use of the rdc of mobile vectors are reported later. Of course, in order to test the applicability of the present strategy, data affected by sizable

mobility need to be discarded.

Even after removing the rdc of the identified mobile amide protons, however, the Q_{free} values calculated using the crystal structure and the tensor obtained from the fit of N-terminal domain pcs remain quite large (0.73-0.52-0.53 for Tb-Tm-Yb DAPk and 0.47-0.31-0.53 for Tb-Tm-Dy DRP-1). We conclude that the disagreement between experimental and calculated rdc values (Figure 5E,F) originates from structural differences. It remains to be assessed whether these differences are only local, i.e. they can be eliminated by slight refinement of the orientations of the NH vectors, or they also reflect more global differences between the structures in the crystalline state and in solution. This is done by two different calculations: one in which the three rdc-derived orientation tensors are fitted to the crystal structure for each domain separately, and another in which the structure of each domain and its three orientation tensors are left free to refine separately taking the crystal structure as a starting point, as one would do for rdc arising from external orienting media.

Separating local from global structural differences

The reduced sets of rdc values of the two domains, considered separately, have been fitted to the crystal structure of the protein using Eq. (2), and the fit shows a significant improvement in the agreement between calculated and observed values with respect to that previously shown (Q_{free} values of 0.21-0.20-0.40 and 0.25-0.27-0.33 for N-terminal domain and C-terminal domain rdc of Tb-Tm-Yb DAPk peptide, respectively; Q_{free} values of 0.23-0.13-0.20 and 0.29-0.16-0.26 for N-terminal domain and C-terminal domain rdc of Tb-Tm-Dy DRP-1 peptide, respectively), although some values are not well fitted (Figure 8A-D). The anisotropies of the orienting tensors for the two domains result to be quite similar, as expected from the fact that the rdc of the two domains span the same range of values. The magnetic susceptibility anisotropies, $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, calculated from these orienting tensors (see Table 2) are similar to those obtained from pcs (although somewhat smaller). The fact that both the N-terminal and the C-terminal domains provide similar magnetic susceptibility anisotropy values confirms that inter-domain reorientational freedom is essentially abolished after binding of CaM to the DAPk/DRP-1 peptides. Importantly, the main directions of the orienting tensors calculated from the rdc of the two domains considered separately are clearly not coincident, as shown in Table 2, but form angles of 20° or larger for at least one of the *x* or *z* axis. Consistently, the simultaneous fit of the same rdc values of both domains is largely unsatisfactory (Q_{free} values of 0.49-0.43-0.48 for Tb-Tm-Yb DAPk and 0.36-0.22-0.33 for Tb-Tm-Dy DRP-1, see Figure 8E,F).

With the reduced set of rdc and the backbone dihedral angles obtained from TALOS[58]

(which were checked to be consistent with the crystal structure) a new structure of the N-terminal and C-terminal domains was calculated as a refinement of the domains in the calculated crystal structure, because the latter has been demonstrated to represent a good starting model. The refinement was performed using the program Xplor-NIH,[59] through the routines PARArestraints for Xplor-NIH[41] and the protocol described in the Materials and Methods section. Two sets of three independent tensors were considered, one set to represent the orienting tensors for the three metal ions responsible for the rdc observed for the N-terminal domain and one set to represent the orienting tensors responsible for the rdc of the C-terminal domain. Even if the calculations were performed in the absence of NOEs, no expansion in the volume of the protein domains was observed. The agreement between all calculated and observed restraints was very satisfactory for both N-terminal and C-terminal domain structures, the Q factor for all rdc actually dropping to 0.12 for both the DAPk and the DRP-1 peptides. Analogously to what found from the fits performed on the crystal structures, the main axes of the orienting tensors obtained for the N-terminal domain are different from those obtained from the C-terminal domain when the two domains are oriented by superimposing them to the corresponding domains in the crystal structure. There is a satisfactory agreement between the changes in the orientation of the tensors in the two calculations. These two calculations therefore strongly suggest that a reorientation of the two CaM domains occurs in solution which, although modest, is outside experimental uncertainty.

Interestingly, in the refinement of the N-terminal domain in the previous calculation, the first helix turns out to have a somewhat different orientation with respect to the crystal structure; in the C-terminal domain there were no major deviations in the protein backbone. As from our strategy, this unexpected observation prompted a further refinement of the N-terminal domain structure adding the pcs restraints to the other restraints and using a single set of tensors for both pcs and rdc restraints relative to the N-terminal domain, the anisotropies of which were calculated from pcs and the refined domain structure. The resulting structure was very similar to the previous one, while the three refined tensors differed only modestly from those determined initially from pcs and the crystal structure, so that no further iteration was needed. The agreement with both pcs and rdc values was excellent (Figure 9), as were the Q values for the rdc (0.11 and 0.21 for DAPk and DRP-1, respectively). Finally, the C-terminal domain was refined by imposing the newly determined anisotropies from the N-terminal domain. Even with the three pcs-derived anisotropies refined on the N-terminal domain the agreement between calculated and experimental rdc remained excellent (Figure 9C,D) and the Q values (0.11 and 0.13 for DAPk and DRP-1, respectively) as low as for the independent refinement. The directions of the main axes of the magnetic

susceptibility anisotropy tensors calculated for the two domains again differ when the domains are superimposed to the whole crystal CaM structure; such differences are relatively large in the DAPk adduct and somewhat smaller in the DRP-1 adduct (see Table 3). This difference is much larger than the error which was estimated by performing multiple calculations after removal of one third of the input pcs and rdc data, perturbed with a stochastic error of 1.5 Hz and 0.1 ppm for rdc and pcs, respectively. The indetermination so obtained for the N-terminal domain (using pcs and rdc) on the tensor axis directions was in fact 5° for the z axis of all metals and for the x, y axes of Tm; it increased to 10° for the x, y axes of Tb and Dy. The indetermination obtained for the C-terminal domain (using rdc only) on the tensor axis directions was 5-8° in all cases besides the x and y axes of Yb (when it is much larger due to the very small rhombic anisotropy). Figures S4-S5 show that the rdc calculated for one domain using the tensor obtained for the other domain are actually very different, beyond the error, from the experimental data, when the two domains are superimposed to the crystal structure, thus pointing out that such paramagnetic restraints detect a structural rearrangement from crystal to solution.

These latter results i) confirm that the pcs values provide reliable and robust tensor values and ii) show that the rdc of both domains are in equally good agreement with the pcs-derived anisotropies as they are with the somewhat smaller rdc-derived anisotropies, pointing to the presence of local mobility as a source of error for the rdc-derived tensors, which would be then less reliable to check the simultaneous presence of global domain rearrangements.

Simultaneous refinement of the two domains using pcs-derived tensors

The refined structure of the protein was then calculated including the same restraints, starting from the structures of the singly refined domains, and including a single triplet of pcs-derived magnetic susceptibility tensors, responsible for both pcs and rdc of the N-terminal and C-terminal domains.[33] The anisotropies of these tensors were fixed to the values obtained from the fit of the pcs data to the refined N-terminal domain. An internal dynamics at 200 K and a minimization were performed (second step in the protocol described in the Material and Methods section). The same refined protein structure was basically obtained also starting from the crystal model and using the same restraints and the same protocol previously used for the refinement of the individual domains, with a single triplet of magnetic susceptibility tensors, the anisotropies of which are responsible for both pcs and rdc of the N-terminal and C-terminal domains.[33] A new set of tensors were recalculated as the new N-terminal structure and the overall backbone structure soon converged (Table 4). Both the values and the

direction of the main axes for the metal susceptibility tensors were in agreement with expectations.[60]

The resulting structures, deposited in PDB (2K0J and 2K61), have a backbone RMSD with the crystal structure of 2.0 and 1.1 Å (in the 6-146 residue range) for the complexes with the DAPk and DRP-1 peptides, respectively, indicating that the solution structures are still relatively close to the crystal structure (see Figure 10). Nevertheless, slight structural differences with respect to the latter are pinpointed, namely in the orientation of the first helix of the N-terminal domain and of the whole C-terminal domain. This very same movement of helix 1 was observed also by Bax et al. in free CaM using rdc and external orienting media.[4] A difference in the relative orientation of the N-terminal and C-terminal domains is actually obtained, particularly in the DAPk adduct, as evidenced by the superimposition of residues 25-65 of the N-terminal domain of the solution structure to the corresponding residues of the crystal structure (Figure 10C,D). In addition, the C-terminal domain, besides the different orientation with respect to the crystal structure, also shows a translational component. The latter obviously cannot be originated by rdc restraints, but arises from pcs. As such, its significance is only qualitative, as in this particular case the pcs values experienced by the C-terminal domain are rather small. Also, the different extent of this translational component, larger in the DAPk and smaller in the DRP-1 complex, should not be taken quantitatively. In any case, a better agreement of the pcs values was also observed for the DAPk complex, the one showing the larger change with respect to the crystal structure (Figure S6). All rdc values are in very good agreement with respect to the refined solution structures (Figure 11), thus indicating that the three sets of data, measured using different lanthanides, are consistent with an unique structural change with respect to the solid state. In conclusion, the paramagnetism-based restraints indicate that the solution structures are similar, but not identical, to the crystal structures. Structural differences with respect to the crystal structure are pinpointed, that can be ascribed to a structural rearrangement from solid state to solution. Similar rearrangements, although of somewhat different amounts, are obtained also including in the calculation only subclasses of pcs/rdc restraints, indicating that all of them point toward the same structural differences.

In the calculations, the peptides were included together with the CaM crystal model, but no restraints have been used, as well as no restraints are provided involving side-chain nuclei. We only refined the crystal structures (where the peptides are present) at low temperature, so that they are kept in place by the Van der Waals contacts only. The same calculations were also repeated without including the peptides, and the same refined CaM structures were obtained. The same structural rearrangement of the protein domains was also

basically calculated when TALOS restraints are not included in the calculations. On the contrary, when rdc are removed, all other restraints (TALOS and/or pcs) do not show any significant rearrangement, in agreement with the fact that they are already in agreement with the crystal structure.[61]

The refined solution structure of CaM when bound to the DAPk peptide was also calculated by adding the C^{α}-H^{α} rdc measured for the Yb and Tm CaM derivatives. The obtained structure is in very good agreement with that calculated without such restraints, and the corresponding rdc values are in excellent agreement with the experimental ones (see Supporting Information, Fig. S7).

The quality of the structures can be monitored using the Q factor[62] calculated from the rdc, which decreases from 0.47 and 0.32, as calculated using the crystal structures of the adduct with DAPk and DRP-1, respectively, and the orienting tensors in best agreement with the crystal structures, to 0.14 and 0.18, as calculated from the refined solution structures of the adduct with DAPk and DRP-1, respectively, and the anisotropy susceptibility tensors calculated for the solution structures. A validation of the structure calculated in the presence of the DAPk peptide has been performed by comparing the rdc measured for $DyCa_3CaM$, which were not used in the structure calculations.[4] The Q_{free} factor[62] calculated from the Dy rdc of non mobile residues decreases from 0.65, as calculated using the crystal structure, to 0.25, as calculated from the refined solution structure, thus indicating an increase in the accuracy of the backbone structure with respect to the crystal structure.[63] A validation of the calculated structure was also performed by comparing the structures obtained using only 2 of the 3 sets of pcs and rdc data.[4] Such structures have a backbone RMSD to the mean of 0.6 and 0.4 Å for the complex in the presence of the DAPk and DRP-1 peptides, respectively (residue range 6-146, see Supporting Information, Fig. S2). When calculating the structures by leaving out either one of the two less numerous of the three sets of pcs and rdc data, the Q_{free} factors for the excluded set of rdc data decrease from 0.54±0.12 to 0.36±0.08, thus indicating an increased backbone accuracy with respect to the crystal structure.[63] The quality of the solution structures is similar to that of the crystal structures, as can be estimated by using PROCHECK_NMR (see Table S3 in the Supporting Information).[64]

It is known that rdc are averaged on time scales larger than that of relaxation data, and therefore more rdc than warranted may be retained by considering that non-mobile residues are only those not showing mobility effects through relaxation measurements. On the other hand, the global changes in the relative orientation of the two domains are determined by the whole ensemble of the rdc restraints, and we have checked that even if one third of them are randomly removed, similar structures are calculated (data not shown). On the other side, even

if all measured rdc, comprising those of N-H vectors identified as mobile from relaxation measurements, are used in the calculations, the obtained structures do not differ significantly from those already described and shown in Fig. 10, the relative orientation of the protein domains remaining basically the same. Therefore, omitting to remove a few more rdc is not expected to alter the picture significantly.

Ensemble average approaches have been recently introduced to improve the agreement in the fit of the data when needed.[65-70] In our case, the data are fitted reasonably well in the assumption of a unique structure. Therefore, we can conclude that no sizable motion occurs in this case between the two CaM domain, and ensemble averages were not attempted. On the contrary, ensemble average approaches[65-70] would be needed to correctly analyse the information on the conformational freedom contained in the rdc of mobile residues. Further studies in this direction are planned by including the rdc of the mobile N-H vectors.

Comparative examination of the crystal and solution structures suggests that the main origin for their difference may lie in the loss of the intermolecular hydrogen bonds which, in the crystalline state, tend to keep helix 1 apart from the rest of the N-terminal domain (see Figure 4). As already shown, intermolecular H-bonds are actually present in the crystal state only between the first helix of the N-terminal domain and the third or fourth helix of the N-terminal domain of a neighboring molecule (E6-D50, E6-N53, E6-R74, E7-T44, E7-D50). When these intermolecular interactions are lost in solution, helix 1 moves closer to the rest of the N-terminal domain (as already observed in free CaM[4]). The occurrence of salt bridges observed in the crystal structure between the N-terminal portion of the bound peptide and both the first helix of the N-terminal domain of CaM and glutamates in the C-terminal domain of CaM may then facilitate the change in orientation of the C-terminal domain of CaM, observed in the solid state with respect to the solution structure.

Concluding remarks

In summary, we have shown that the simultaneous use of paramagnetic restraints such as pcs-derived orientation tensor and self-orientation rdc enhances the possibility to refine solution structures starting from crystal structures in multidomain proteins, and that the resulting structures have the same quality as the crystal structures and provide a more accurate description of the protein structure in solution.

Residual dipolar couplings originating from external orienting media were previously used to detect structural rearrangements in solution with respect to the crystal structure.[5;40] For a two-domain protein such a CaM, paramagnetic rdc offer the advantage that i) relative conformational freedom can be immediately assessed and, if present, dealt with;[9;10] ii) in the absence of significant conformational freedom, the global orientation tensor can be

independently and precisely determined from pcs and iii) the *relative* rearrangement of a domain or a secondary structure element with respect to the metal-bearing domain can be more accurately detected.

Although in the cases selected here the protein has an intrinsic paramagnetic metal binding site,[9;10] in other cases paramagnetic tags can be used,[11;71-73] provided they are rigidly attached to the protein.[56-59] This approach is particularly useful for proteins with domains experiencing flexibility, such as multidomain proteins, and for protein-protein adducts. However, all the more so, the approach can be applied to refine single domain proteins containing a metal ion binding site, either natural or artificial and is, therefore, quite general.

Acknowlegements

We thank Mr. Paul Ionescu for providing some apo-calmodulin samples, and Inaki de Diego for discussions. This work has been supported by Ente Cassa di Risparmio, MIUR-FIRB contracts RBLA032ZM7 and RBIP06LSS2, and by European Commission, contracts EU-NMR 026145, SPINE2-COMPLEXES 031220, and LSHG-CT-2004-512052.

Reference List

(1) Fragai, M.; Luchinat, C.; Parigi, G. *Acc.Chem.Res.* **2006**, *39*, 909-917.

(2) Clore, G. M.; Gronenborn, A. M. *Proc.Natl.Acad.Sci.USA* **1998**, *95*, 5891-5898.

(3) Spronk, C. A. E. M.; Nabuurs, S. B.; Krieger, E.; Vriend, G.; Vuister, G. W. *Prog.NMR Spectrosc.* **2004**, *45*, 315-337.

(4) Chou, J. J.; Li, S.; Klee, C. B.; Bax, A. *Nature Struct.Biol.* **2001**, *8*, 990-997.

(5) Skrynnikov, N. R.; Goto, N. K.; Yang, D.; Choy, W.-Y.; Tolman, J. R.; Mueller, G. A.; Kay, L. E. *J.Mol.Biol.* **2000**, *295*, 1265-1273.

(6) Goto, N. K.; Skrynnikov, N. R.; Dahlquist, F. W.; Kay, L. E. *J.Mol.Biol.* **2001**, *308*, 745-764.

(7) Gochin, M.; Roder, H. *Protein Sci.* **1995**, *4*, 296-305.

(8) Fowler, B. A.; Tian, F.; Al-Hashimi, H. M.; Prestegard, J. H. *J.Mol.Biol.* **2000**, *304*, 447-460.

(9) Bertini, I.; Del Bianco, C.; Gelis, I.; Katsaros, N.; Luchinat, C.; Parigi, G.; Peana, M.; Provenzani, A.; Zoroddu, M. A. *Proc.Natl.Acad.Sci.USA* **2004**, *101*, 6841-6846.

(10) Bertini, I.; Gupta, Y. K.; Luchinat, C.; Parigi, G.; Peana, M.; Sgheri, L.; Yuan, J. *J.Am.Chem.Soc.* **2007**, *129*, 12786-12794.

(11) Pintacuda, G.; Park, A. Y.; Keniry, M. A.; Dixonj, N. E.; Otting, G. *J.Am.Chem.Soc.* **2006**, *128*, 3696-3702.

(12) Tolman, J. R.; Al-Hashimi, H. M.; Kay, L. E.; Prestegard, J. H. *J.Am.Chem.Soc.* **2001** , *123*, 1416-1424.

(13) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Bruschweiler, R. *J.Am.Chem.Soc.* **2001**, *123*, 6098-6107.

(14) Tolman, J. R. *J.Am.Chem.Soc.* **2002**, *124*, 12020-12030.

(15) Bertini, I.; Luchinat, C.; Parigi, G. *Concepts Magn.Reson.* **2002**, *14*, 259-286.

(16) Baber, J. L.; Szabo, A.; Tjandra, N. *J.Am.Chem.Soc.* **2001**, *123*, 3953-3959.

(17) Barbato, G.; Ikura, M.; Kay, L. E.; Pastor, R. W.; Bax, A. *Biochemistry* **1992**, *31*, 5269-5278.

(18) Cohen, O.; Feinstein, E.; Kimchi, A. *The EMBO Journal* **1997**, *16*, 998-1008.

(19) Hoeflich, K. P.; Ikura, M. *Cell* **2002**, *108*, 739-742.

(20) Elshorst, B.; Hennig, M.; Forsterling, H.; Diener, A.; Maurer, M.; Schulte, P.; Schwalbe, H.; Griesinger, C.; Krebs, J. F.; Schmid, H.; Carafoli, E. *Biochemistry* **1999**, *38*, 12320-12332.

(21) Majava, V.; Petoukhov, M. V.; Hayashi, N.; Pirilä, P.; Svergun, D. I.; Kursula, P. *BMC.Struct.Biol.* **2008**, *8*, 10.

(22) Bertini, I.; Gelis, I.; Katsaros, N.; Luchinat, C.; Provenzani, A. *Biochemistry* **2003**, *42*, 8011-8021.

(23) Kabsch, W. *J.Appl.Crystallogr.* **1993**, *26*, 795-800.

(24) Kursula, P. *J.Appl.Crystallogr.* **2004**, *37*, 347-348.

(25) Vagin, A.; Teplyakov, A. *J.Appl.Crystallogr.* **1997**, *30*, 1022-1025.

(26) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Cryst.* **1997**, *D53*, 240-255.

(27) Winn, M. D.; Isupov, M. N.; Murshudov, G. N. *Acta Cryst.D* **2001**, *57*, 122-133.

(28) Jones, T. A.; Zou, J. Y.; Cowtan, S. W.; Kjeldgaard, M. *Acta Cryst.* **1991**, *A47*, 110-119.

(29) Perrakis, A.; Morris, R. J. H.; Lamzin, V. S. *Nat Struct Biol* **1999**, *6*, 458-463.

(30) Ottiger, M.; Delaglio, F.; Bax, A. *J.Magn.Reson.* **1998**, *131*, 373-378.

(31) Kay, L. E.; Torchia, D. A.; Bax, A. *Biochemistry* **1989**, *28* , 8972-8979.

(32) Goddard, T. D. and Kneller, D. G. SPARKY 3, University of California, San Francisco. 2000.

(33) Bertini, I.; Luchinat, C.; Parigi, G. *Progr.NMR Spectrosc.* **2002**, *40*, 249-273.

(34) Banci, L.; Bertini, I.; Huber, J. G.; Luchinat, C.; Rosato, A. *J.Am.Chem.Soc.* **1998**, *120*, 12903-12909.

(35) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc.Natl.Acad.Sci.USA* **1995**, *92*, 9279-9283.

(36) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Nature Struct.Biol.* **1997**, *4*, 292-297.

(37) Bax, A.; Tjandra, N. *Nature Struct.Biol.* **1997**, *4*, 254-256.

(38) Bothner-By, A. A.; Domaille, J. P.; Gayathri, C. *J.Am.Chem.Soc.* **1981**, *103*, 5602-5603.

(39) Bertini, I.; Luchinat, C.; Parigi, G.; Pierattelli, R. *ChemBioChem* **2005**, *6*, 1536-1549.

(40) Chou, J. J.; Li, S.; Bax, A. *J Biomol NMR* **2000**, *18*, 217-227.

(41) Banci, L.; Bertini, I.; Cavallaro, G.; Giachetti, A.; Luchinat, C.; Parigi, G. *J.Biomol.NMR* **2004**, *28*, 249-261.

(42) Ikura, M.; Clore, G. M.; Gronenborn, A. M.; Zhu, G.; Clee, C.; Bax, A. *Science* **1992** , *256*, 632-638.

(43) Meador, W. E.; Means, A. R.; Quiocho, F. A. *Science* **1992**, *257*, 1251-1255.

(44) Meador, W. E.; Means, A. R.; Quiocho, F. A. *Science* **1993**, *262*, 1718-1721.

(45) Maximciuc, A. A.; Putkey, J. A.; Shamoo, Y.; MacKenzie, K. R. *Structure* **2006**, *14*, 1547-1556.

(46) Schumacher, M. A.; Rivard, A. F.; Bächinger, H. P.; Adelman, J. P. *Nature* **2001**, *410*, 1120-1124.

(47) Aoyagi, M.; Arvai, A. S.; Tainer, J. A.; Getzoff, E. D. *EMBO J.* **2003**, *22*, 766-775.

(48) Liu, M.; Chen, T. Y.; Ahamed, B.; Li, J.; Yau, K. W. *Science* **1994**, *266*, 1348-1354.

(49) Al-Hashimi, H. M.; Valafar, H.; Terrell, M.; Zartler, E. R.; Eidsness, M. K.; Prestegard, J. H. *J.Magn.Reson.* **2000**, *143*, 402-406.

(50) John, M.; Otting, G. *ChemPhysChem* **2007**, *8*, 2309-2313.

(51) Balayssac, S.; Bertini, I.; Luchinat, C.; Parigi, G.; Piccioli, M. *J.Am.Chem.Soc.* **2006**, *128*, 15042-15043.

(52) When the pcs of the, N, C^{α}, C^{β}, C' nuclei are included, the calculated tensors remain basically the same (the anisotropy values changing up to a maximum of 4%).

(53) Frederick, K. K.; Kranz, J. K.; Wand, A. J. *Biochemistry* **2006**, *45*, 9841-9848.

(54) Wang, T.; Frederick, K. K.; Igumenova, T. I.; Wand, A. J.; Zuiderweg, E. R. P. *J.Am.Chem.Soc.* **2005**, *127*, 828-829.

(55) Marlow, M. S.; Wand, A. J. *Biochemistry* **2006**, *45*, 8732-8741.

(56) de la Torre, J. G.; Huertas, M. L.; Carrasco, B. *J.Magn.Reson.* **2000**, *147*, 138-146.

(57) Larsson, G.; Martinez, G.; Schleucher, J.; Wijmenga, S. S. *J Biomol NMR* **2004**, *27*, 291-312.

(58) Cornilescu, G.; Delaglio, F.; Bax, A. *J Biomol NMR* **1999**, *13*, 289-302.

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

(60) Bertini, I.; Donaire, A.; Jiménez, B.; Luchinat, C.; Parigi, G.; Piccioli, M.; Poggi, L. *J.Biomol.NMR* **2001**, *21*, 85-98.

(61) A restrained energy minimization was also attempted while keeping the position of all C^{α} atoms fixed to the crystal coordinates, in order to check whether a good agreement of the rdc data could be obtained without significant changes in the protein backbone structures. The calculations actually indicated that structures of sizably larger energy are obtained, thus pointing out that changes in the backbone protein structure are indeed needed.

(62) Cornilescu, G.; Marquardt, J.; Ottiger, M.; Bax, A. *J.Am.Chem.Soc.* **1998**, *120*, 6836-6837.

(63) Clore, G. M.; Kuszewski, J. *J.Am.Chem.Soc.* **2003**, *125*, 1518-1525.

(64) Laskowski, R. A.; Rullmann, J. A. C.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. *J.Biomol.NMR* **1996**, *8*, 477-486.

(65) Iwahara, J.; Schwieters, C. D.; Clore, G. M. *J.Am.Chem.Soc.* **2004**, *126*, 5879-5896.

(66) Clore, G. M.; Schwieters, C. D. *J.Am.Chem.Soc.* **2004**, *126*, 2923-2938.

(67) Clore, G. M.; Schwieters, C. D. *J.Mol.Biol.* **2006**, *355*, 879-886.

(68) Gsponer, J.; Hopearuoho, H.; Whittaker, S. B.-M.; Spence, G. R.; Moore, G. R.; Paci, E.; Radford, S. E.; Vendruscolo, M. *Proc.Natl.Acad.Sci.U.S.A* **2006**, *103*, 99-104.

(69) Lange, O. F.; Lakomek, N.-A.; Farès, C.; Schröder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320*, 1471-1475.

(70) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128-132.

(71) Wöhnert, J.; Franz, K. J.; Nitz, M.; Imperiali, B.; Schwalbe, H. *J.Am.Chem.Soc.* **2003**, *125*, 13338-13339.

(72) Prudencio, M.; Rohovec, J.; Peters, J. A.; Tocheva, E.; Boulanger, M. J.; Murphy, M. E.; Hupkes, H. J.; Koster, W.; Impagliazzo, A.; Ubbink, M. *Chemistry - A European Journal* **2004**, *5*, 3252-3260.

(73) Ikegami, T.; Verdier, L.; Sakhaii, P.; Grimme, S.; Pescatore, P.; Saxena, K.; Fiebig, K. M.; Griesinger, C. *J.Biomol.NMR* **2004**, *29*, 339-349

Table 1. H-bonds, salt bridges and Van der Waals contacts between the CaM-DAPk peptide complex or the CaM-DRP-1 peptide complex and its neighbours in the crystalline state.

Mol.No	Residue	Atom	Mol.No.	Residue	Atom	Distance (Å)	Angle (°)
CaM-DAPk peptide							
H-bonds							
1	3 GLN	HE22	7	78 ASP	OD1	2.33	9.1
1	6 GLU	HN	3	50 ASP	OD2	1.84	18.9
1	7 GLU	HN	3	50 ASP	OD1	1.87	23.3
1	44 THR	HG1	5	7 GLU	OE2	2.01	9.1
1	50 ASP	HD2	5	6 GLU	N	1.83	8.7
1	53 ASN	HD22	5	6 GLU	OE1	1.86	11.6
1	74 ARG+	HE	7	6 GLU	OE1	1.59	6.3
3	44 THR	HG1	1	7 GLU	OE2	2.01	9.1
3	50 ASP	HD2	1	6 GLU	N	1.83	8.7
3	53 ASN	HD22	1	6 GLU	OE1	1.86	11.6
5	6 GLU	HN	1	50 ASP	OD2	1.84	18.9
5	7 GLU	HN	1	50 ASP	OD1	1.87	23.3
7	3 GLN	HE22	1	78 ASP	OD1	2.33	9.1
7	74 ARG+	HE	1	6 GLU	OE1	1.59	6.3
Salt bridges							
1	74 ARG+	HH22	7	6 GLU	OE2	2.01	7.4
7	74 ARG+	HH22	1	6 GLU	OE2	2.01	7.4
VdW contacts							
1	6 GLU	HN	3	50 ASP	HD2		
1	6 GLU	HE2	3	49 GLN	HG2		
1	6 GLU	HE2	7	74 ARG+	HH22		
1	22 ASP	O	2	148 LYS+	HG3		
1	30 LYS+	HG3	2	79 THR	HG21		
1	49 GLN	HG2	5	6 GLU	HE2		
1	50 ASP	HD2	5	6 GLU	HN		
1	60 ASN	HB3	4	320 ARG+	HH11		
1	74 ARG+	HH22	7	6 GLU	HE2		
1	79 THR	HG21	11	30 LYS+	HG3		
1	148 LYS+	HG3	11	22 ASP	O		
1	320 ARG+	HH11	8	60 ASN	HB3		
CaM-DRP-1 peptide							
H-bonds							
1	6 GLU	HN	3	50 ASP	OD2	1.77	15.4
1	7 GLU	HN	3	50 ASP	OD1	2.29	17.5
1	50 ASP	HD2	5	6 GLU	N	1.76	1.4
1	53 ASN	HD22	5	6 GLU	OE1	1.82	8.5
1	74 ARG+	HE	7	6 GLU	OE1	1.75	18.7
1	78 ASP	HD2	12	53 ASN	O	1.67	28.2
3	50 ASP	HD2	1	6 GLU	N	1.76	1.4
3	53 ASN	HD22	1	6 GLU	OE1	1.82	8.5
7	74 ARG+	HE	1	6 GLU	OE1	1.75	18.7
5	6 GLU	HN	1	50 ASP	OD2	1.77	15.4
5	7 GLU	HN	1	50 ASP	OD1	2.29	17.5
14	78 ASP	HD2	1	53 ASN	O	1.67	28.2
Salt bridges							
1	74 ARG+	HH22	7	6 GLU	OE2	1.97	12.3
7	74 ARG+	HH22	1	6 GLU	OE2	1.97	12.3
VdW contacts							
1	6 GLU	HN	3	50 ASP	HD2		
1	6 GLU	HE2	3	49 GLN	HG2		
1	6 GLU	HE2	7	74 ARG+	HH22		
1	7 GLU	HG3	3	47 GLU	HG3		
1	47 GLU	HG3	5	7 GLU	HG3		
1	49 GLN	HG2	5	6 GLU	HE2		

1	50 ASP HD2	5	6 GLU HN
1	53 ASN O	14	78 ASP OD2
1	74 ARG+ HH22	7	6 GLU HE2
1	78 ASP OD2	12	53 ASN O
1	101 SER HG	2	131 ASP HB2
1	131 ASP HB2	16	101 SER HG

Table 2 Magnetic susceptibility anisotropy values calculated by fitting the experimental data to the crystal structure and angles between the *z* and *x* axes of the magnetic susceptibility anisotropy tensors

DAPk

	Tb^{3+}	Tm^{3+}	Yb^{3+}	z^{Tb}-x^{Tm}	x^{Tb}-z^{Tm}	z^{Tb}-x^{Yb}	x^{Tb}-z^{Yb}					
- from N-terminal domain pcs												
$\Delta\chi_{ax}$	35.5	25.1	8.68×10^{-32} m^3	6°	26°	24°	7°					
$\Delta\chi_{rh}$	−16.9	−11.0	-1.18×10^{-32} m^3									
				Δz^{Tb}	Δx^{Tb}	Δz^{Tm}	Δx^{Tm}	Δz^{Yb}	Δx^{Yb}			
- from N-terminal domain rdc of non mobile HN												
D_a	14.0	10.8	4.01 Hz									
R	0.56	0.33	0.04									
$\Delta\chi_{ax}$	31.0	24.0	8.90×10^{-32} m^3	13°	26°	8°	7°	2°	1°	(a)		
$\Delta\chi_{rh}$	−17.4	−7.8	-0.36×10^{-32} m^3									
- from C-terminal domain rdc of non mobile HN												
D_a	13.5	8.43	3.15 Hz									
R	0.43	0.46	0.27									
$\Delta\chi_{ax}$	30.0	18.7	6.98×10^{-32} m^3	31°	24°	16°	35°	22°	24°	(a)		
$\Delta\chi_{rh}$	−12.9	−8.6	-1.90×10^{-32} m^3									

DRP1

	Tb^{3+}	Tm^{3+}	Dy^{3+}	z^{Tb}-x^{Tm}	x^{Tb}-z^{Tm}	z^{Tb}-y^{Dy}	x^{Tb}-z^{Dy}					
- from N-terminal domain pcs												
$\Delta\chi_{ax}$	37.3	22.7	-40.6×10^{-32} m^3	8°	17°	12°	14°					
$\Delta\chi_{rh}$	−14.2	−12.5	19.6×10^{-32} m^3									
				Δz^{Tb}	Δx^{Tb}	Δz^{Tm}	Δx^{Tm}	Δz^{Dy}	Δx^{Dy}			
				10°	8°	3°	3°			(a)		
- from N-terminal domain rdc of non mobile HN												
D_a	11.9	10.0	13.4 Hz									
R	0.75	0.23	0.42									
$\Delta\chi_{ax}$	26.3	22.2	-29.7×10^{-32} m^3	19°	20°	7°	37°	16°	30°	(b)		
$\Delta\chi_{rh}$	−19.8	−5.2	12.5×10^{-32} m^3									
- from C-terminal domain rdc of non mobile HN												
D_a	13.8	9.29	14.5 Hz									
R	0.49	0.40	0.51									
$\Delta\chi_{ax}$	30.7	20.6	-32.1×10^{-32} m^3	7°	11°	2°	9°	3°	6°	(b)		
$\Delta\chi_{rh}$	−15.0	−8.2	16.5×10^{-32} m^3									

(a) angle with the corresponding tensor axis calculated from N-terminal domain pcs for DAPk;
(b) angle with the corresponding tensor axis calculated from N-terminal domain pcs for DRP-1

Table 3. Angles between the corresponding magnetic susceptibility anisotropy tensor axes obtained for the N-terminal and the C-terminal domains, calculated through a separate fit of pcs and rdc data referring to the two domains

DAPk

	Δx	Δy	Δz
Tb	14°	21°	24°
Tm	25°	23°	14°
Yb	16°	2°	17°

DRP1

	Δx	Δy	Δz
Tb	17°	16°	10°
Tm	13°	20°	20°
Dy	11°	9°	8°

Table 4. Magnetic susceptibility anisotropy values obtained from the N-terminal domain pcs and refined solution structures[a]. The uncertainties for the axial component is estimated around 10%, as calculated by removal of different subsets of data. The angles between the z and x axes of the magnetic susceptibility anisotropy tensors are also reported.

DAPk

	Tb^{3+}	Tm^{3+}	Yb^{3+}	z^{Tb}-x^{Tm}	x^{Tb}-z^{Tm}	z^{Tb}-x^{Yb}	x^{Tb}-z^{Yb}
$\Delta\chi_{ax}$	39.4	25.7	8.95×10^{-32} m^3	4°	9°	16°	12°
$\Delta\chi_{rh}$	−15.1	−12.1	-1.27×10^{-32} m^3				

DRP-1

	Tb^{3+}	Tm^{3+}	Dy^{3+}	z^{Tb}-x^{Tm}	x^{Tb}-z^{Tm}	z^{Tb}-y^{Dy}	x^{Tb}-z^{Dy}
$\Delta\chi_{ax}$	40.0	24.4	-40.3×10^{-32} m^3	5°	11°	9°	19°
$\Delta\chi_{rh}$	−17.1	−13.0	-17.7×10^{-32} m^3				

[a]Tensor axes of Tb^{3+} and Tm^{3+} calculated in the presence of the DAPk peptide are all within 3° with respect to those calculated in the presence of the DRP-1 peptide

Figure 1

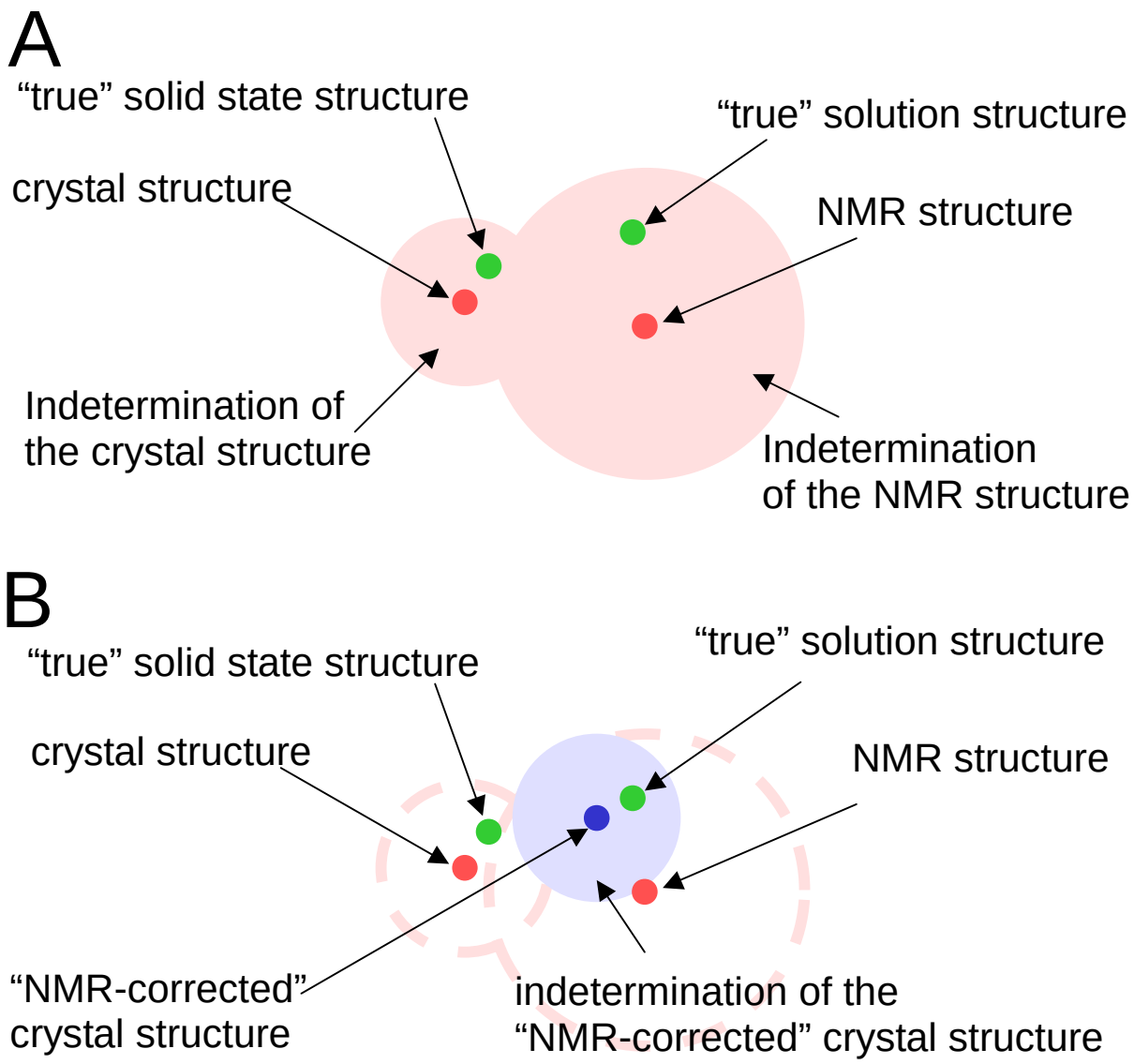

A
“true” solid state structure
crystal structure
“true” solution structure
NMR structure
Indetermination of
the crystal structure
Indetermination
of the NMR structure
B
“true” solid state structure
crystal structure
“true” solution structure
NMR structure
“NMR-corrected”
crystal structure
indetermination of the
“NMR-corrected” crystal structure

Figure 2. Distribution of the rdc for the CaM N-terminal and C-terminal domains measured in the presence of an external orienting device[4] or measured for the derivative obtained substituting Tm^{3+} to Ca^{2+} in the second calcium binding site of the N-terminal domain.[9]

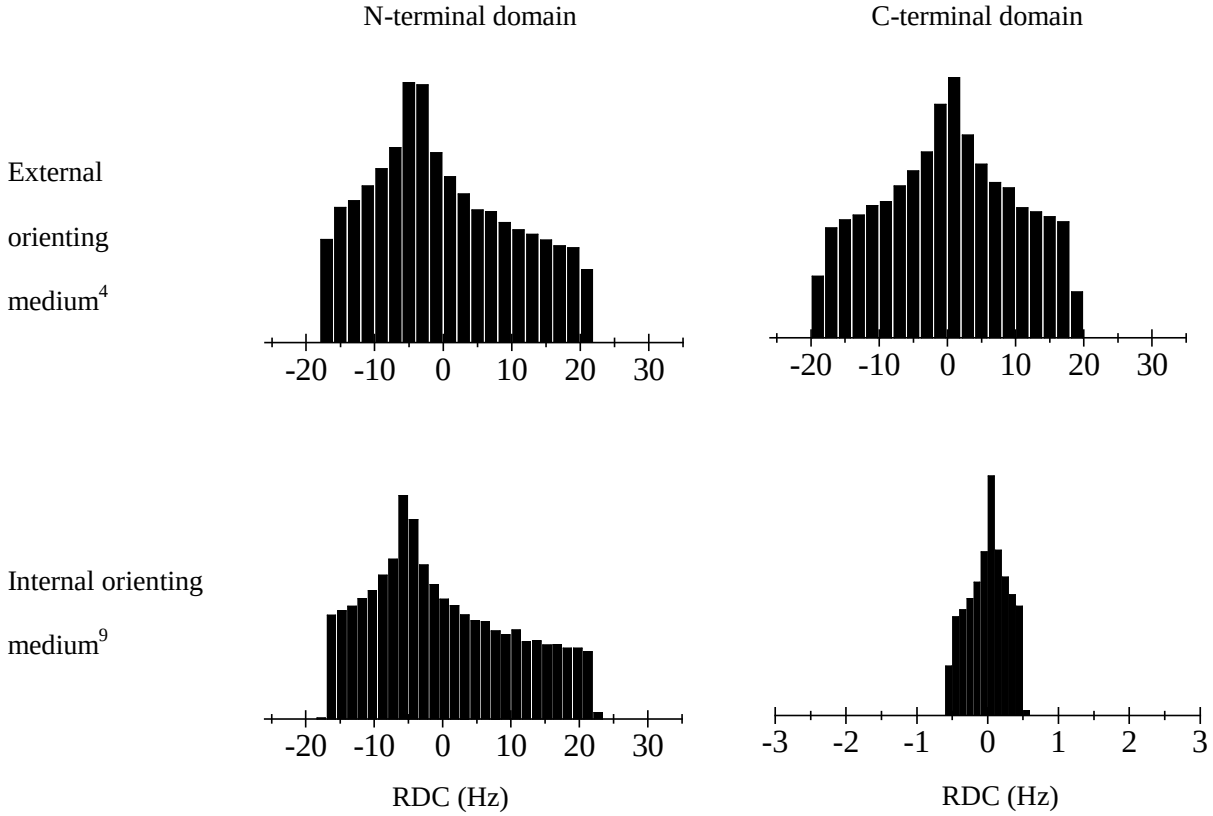


Figure 3 Stereoview of the crystal structures of CaM bound to DAPk (in yellow) or to DRP-1 (in orange) peptides superimposed to the PDB structures 1CDL, 1NIW and 1SY9 (in blue)

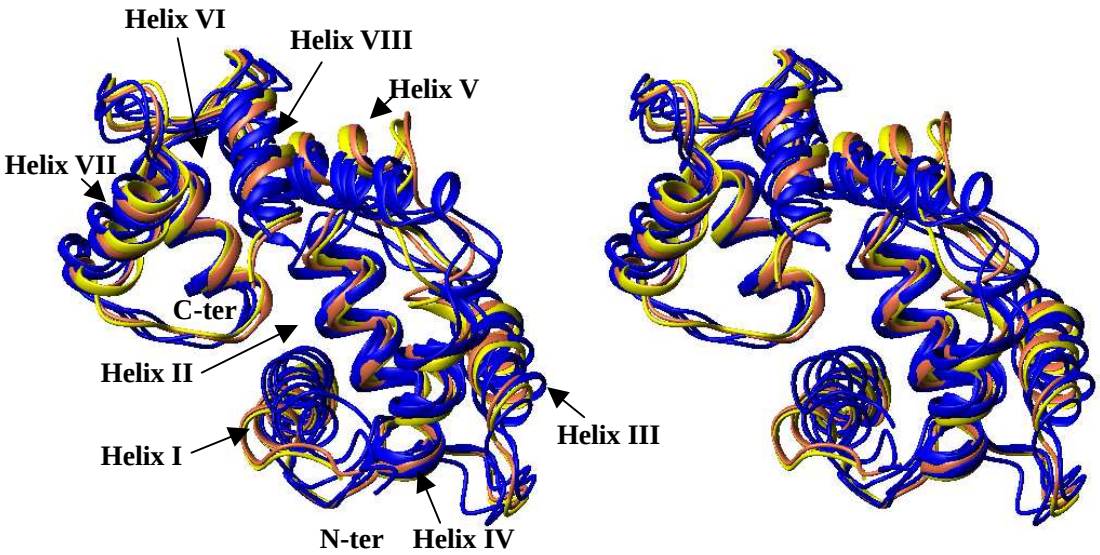
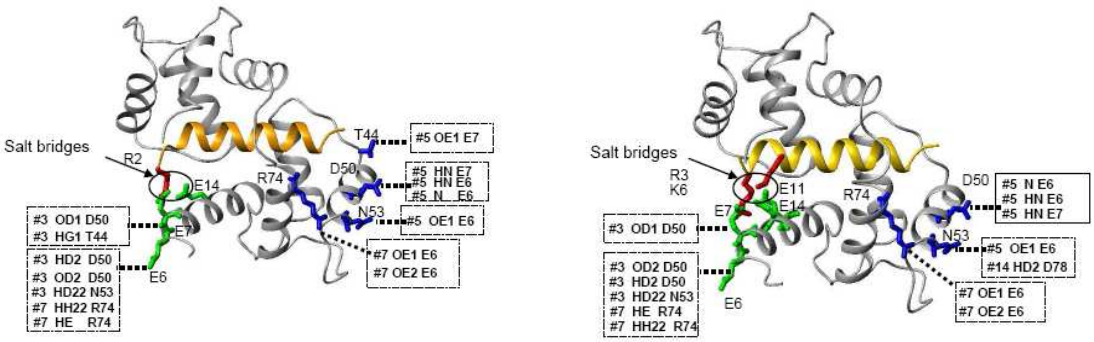


Figure 4. Pattern of H-bonds and salt bridges which are lost or altered on passing from the solid state crystalline form to solution. All the intermolecular H-bonds between the N-terminal domain and neighbor molecules are lost in both complexes.



A) DAPk **B) DRP-1**

Figure 5. Observed versus calculated pcs of N-terminal CaM domain nuclei for the three lanthanide-substituted CaM samples in the adduct with DAPk (A) and DRP-1 (B) peptides. Calculations have been performed using the protein crystal structure. Observed N-terminal and C-terminal CaM domain rdc values versus rdc values calculated using the crystal structure and the paramagnetic susceptibility anisotropy tensor obtained from the fit of the N-terminal domain pcs (grey symbols indicate mobile residues) in the presence of the DAPk (C) or the DRP-1 (D) peptides.

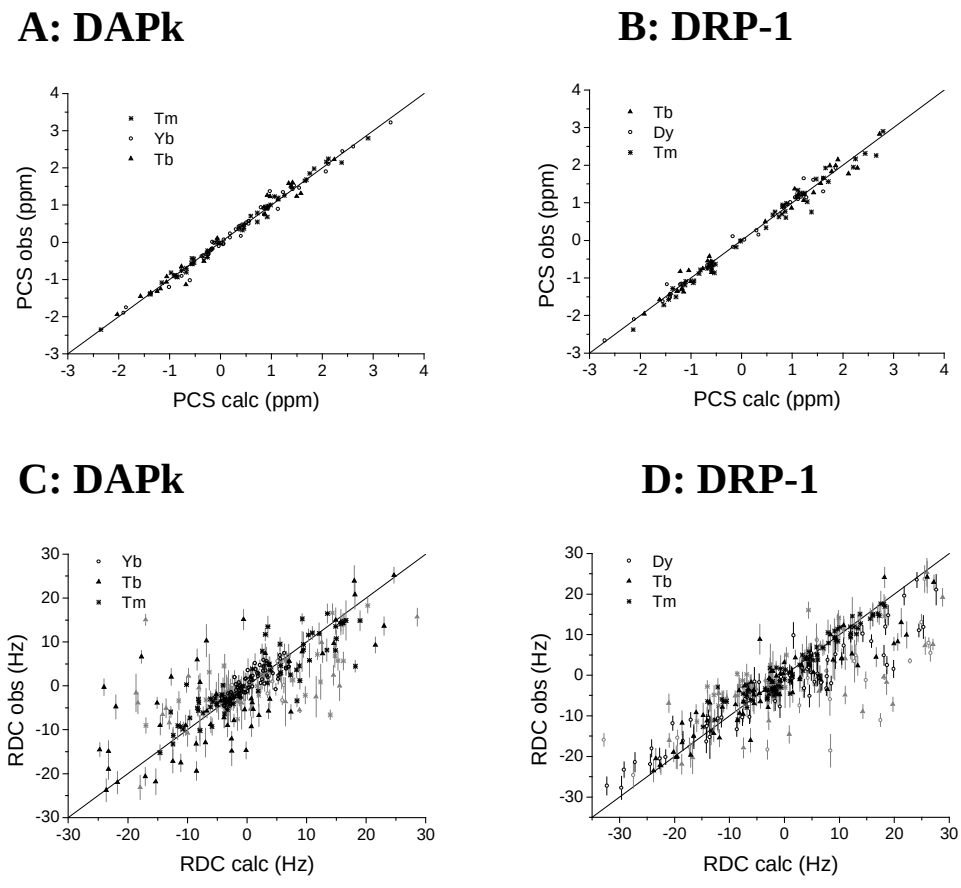


Figure 6. R_1 and R_2 measurements of CaM in the presence of the DAPk peptide performed at 700 MHz of proton Larmor frequency are shown as solid points. HYDRONMR predictions are shown as bars. The upper bars indicate the residues belonging to secondary structural elements.

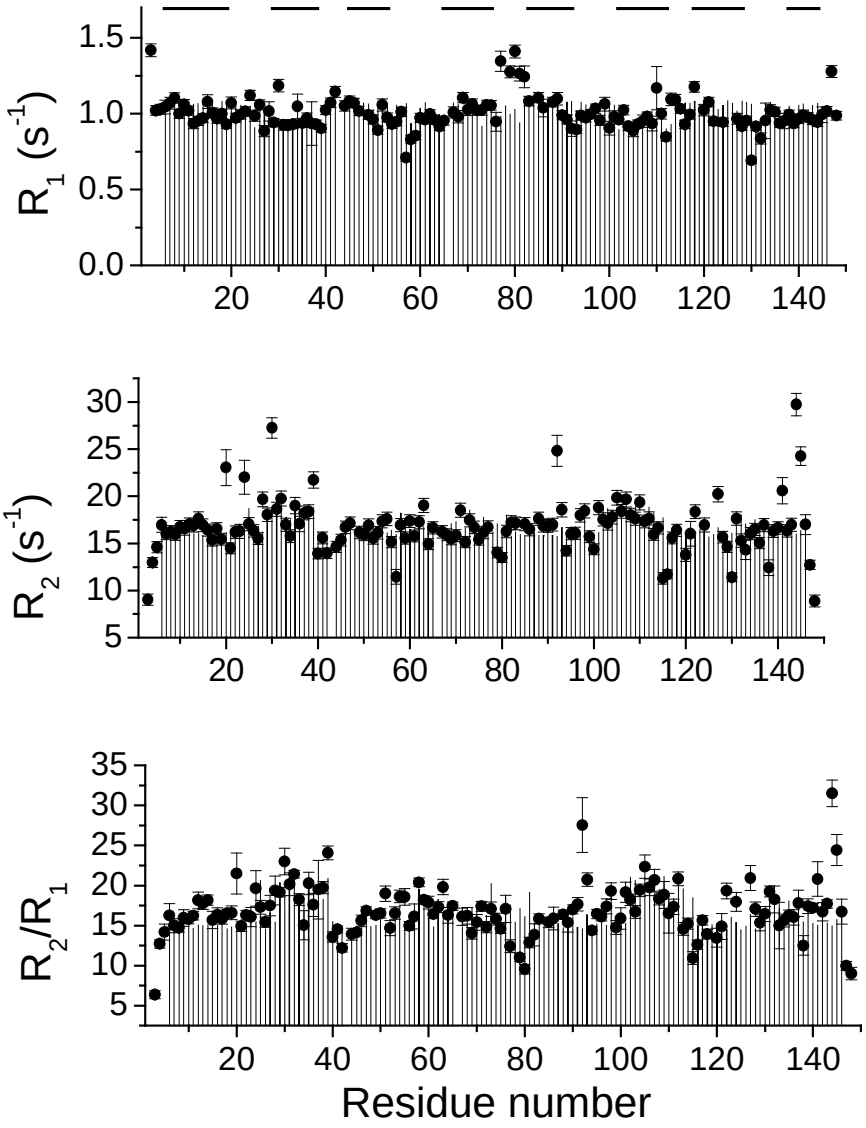


Figure 7. R_1 and R_2 measurements of CaM in the presence of the DRP-1 peptide performed at 700 MHz of proton Larmor frequency are shown as solid points. HYDRONMR predictions are shown as bars. The upper bars indicate the residues belonging to secondary structural elements.

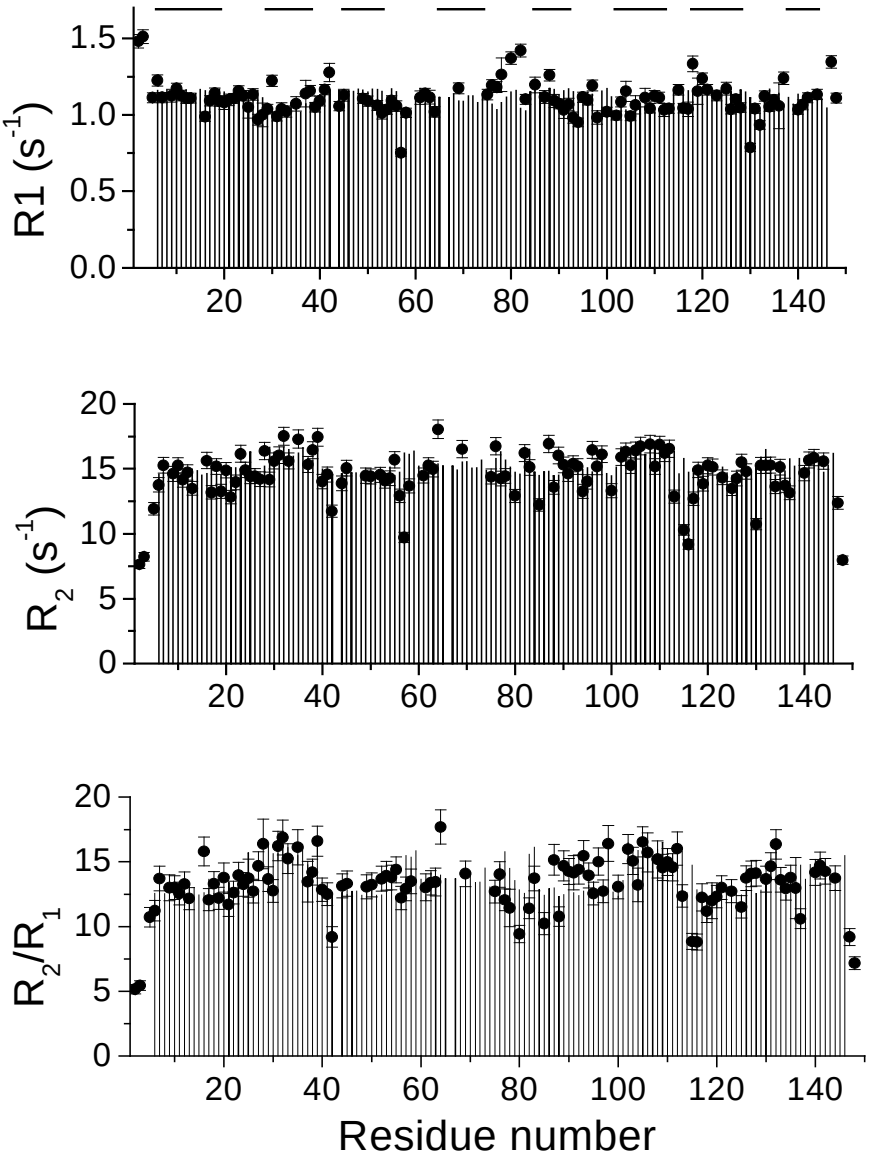


Figure 8. Calculated versus observed rdc of N-terminal and/or C-terminal domain nuclei, for the three lanthanide-substituted CaM samples in the adduct with the DAPk peptide (A, C, E) or the DRP-1 (B, D, F) peptide. Calculations have been performed using the isolated domains (A-D) or the whole protein (E,F) crystal structure.

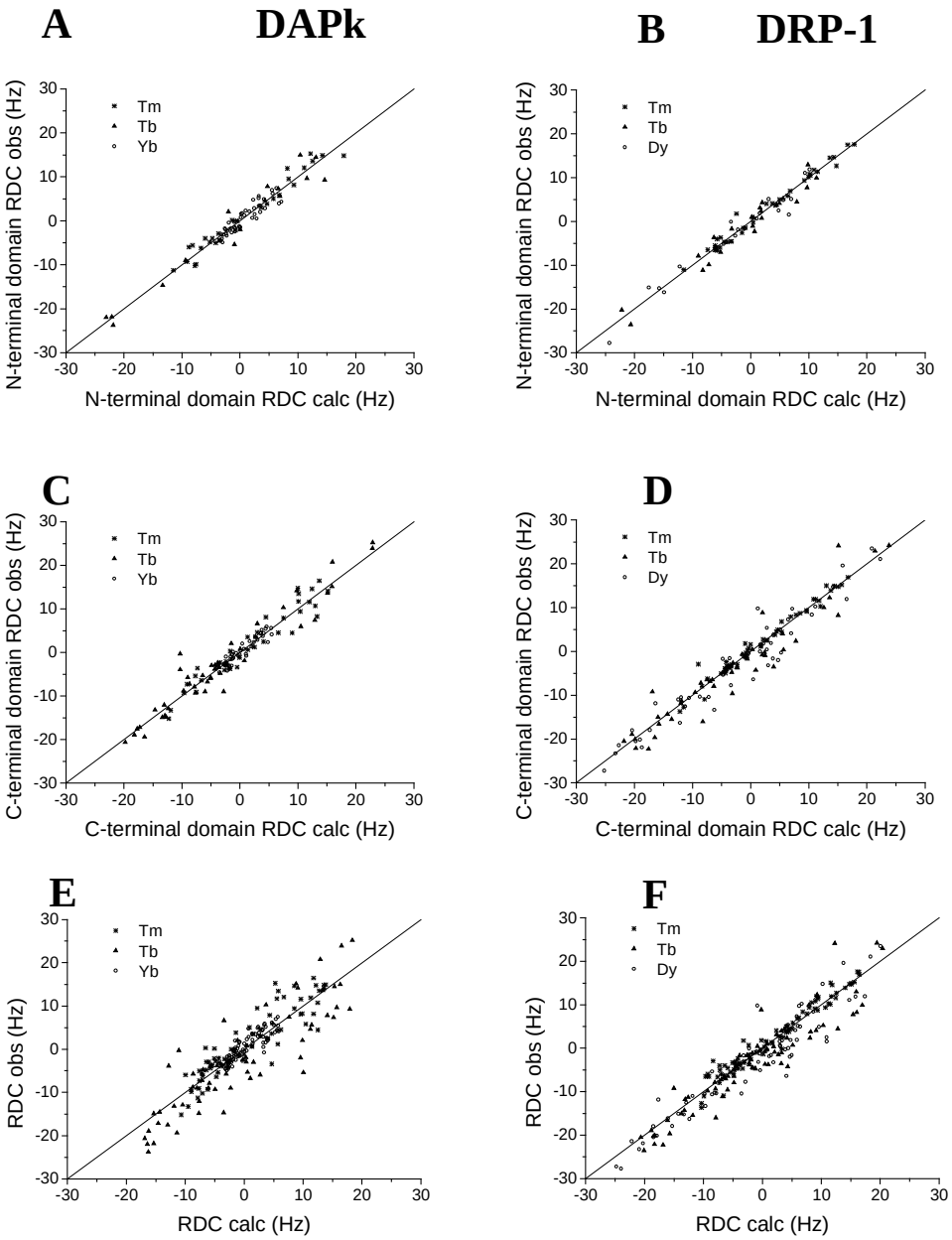


Figure 9. Calculated versus observed pcs and rdc for the three lanthanide-substituted CaM samples in the adduct with the DAPk peptide (A, C) or the DRP-1 (B, D) peptide, using the solution structure calculated by refining the crystal structure with the experimental pcs and rdc and different sets of tensors for the N-terminal and C-terminal domains.

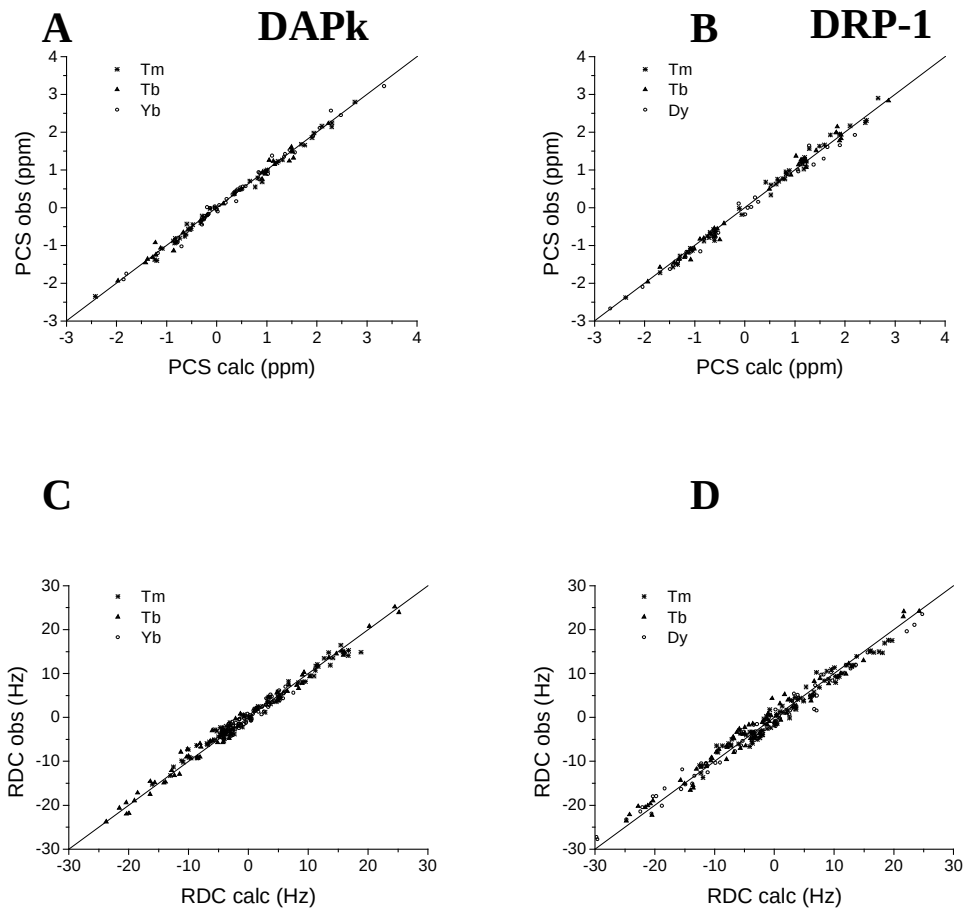


Figure 10. The CaM solution structures (in blue) are shown together with the crystal structures (in yellow/orange) calculated in the presence of the DAPk (A, C) or the DRP-1 peptides (B, D). The structures are shown with fitting the residues in the 6-146 range (A, B) or in the 25-65 range (C, D).

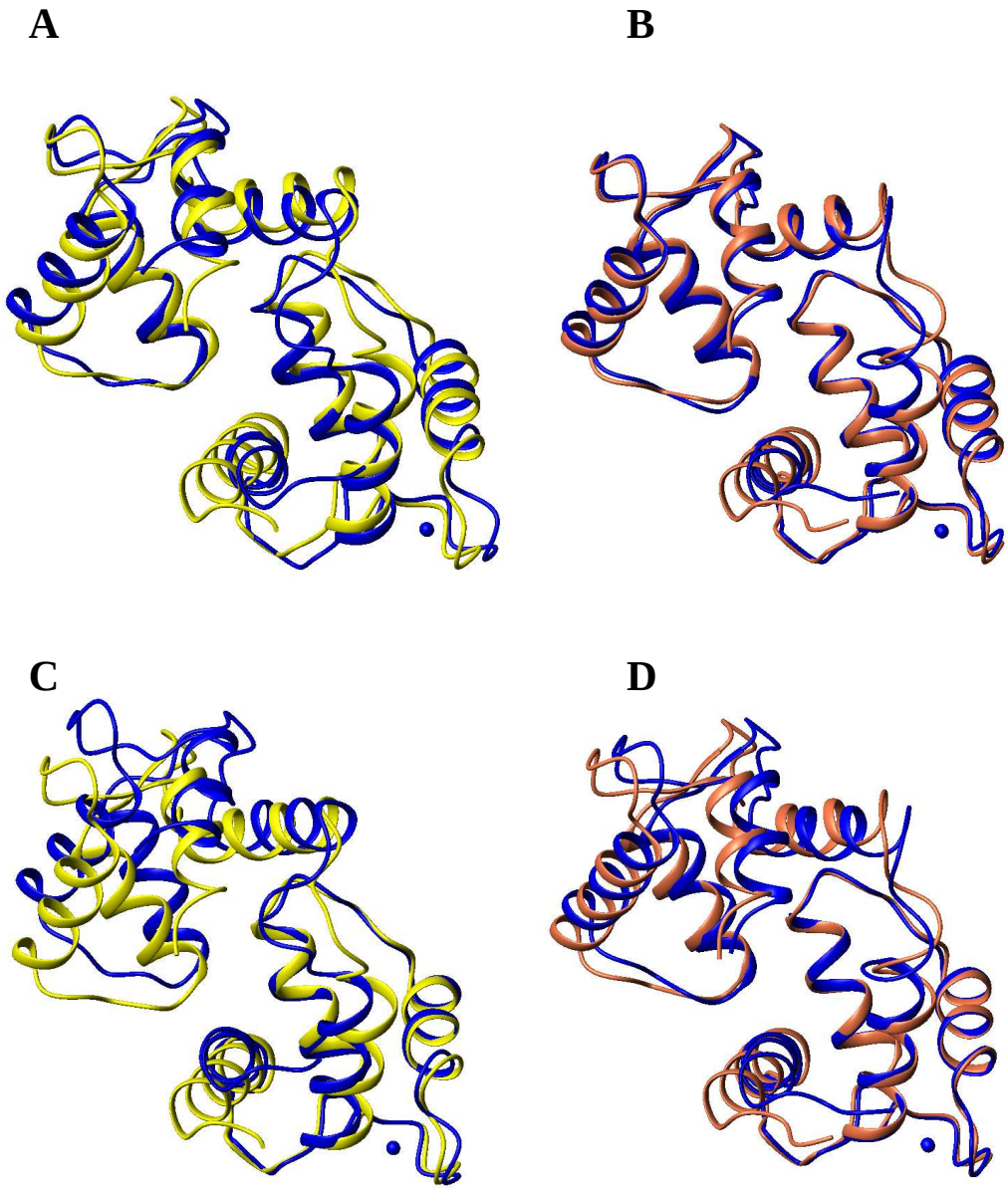
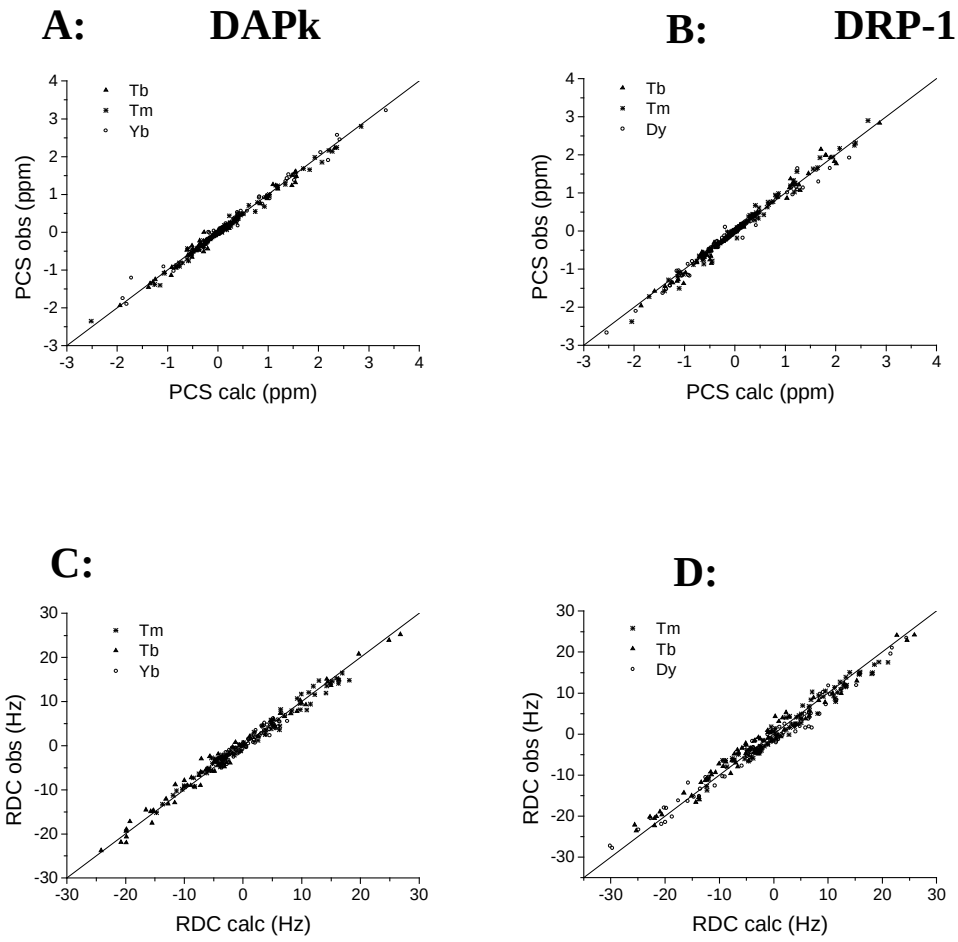


Figure 11. Observed pcs and rdc values versus the values calculated using the refined solution structure for the DAPk (A, C) and the DRP-1 (B, D) peptides.



3.3

Solution structure and dynamics of S100A5 in the apo and Ca^{2+} -bound states

Ivano Bertini,[1,2*] Soumyasri Das Gupta,[1] Xiaoyu Hu,[1] Tilemachos Karavelas,[1,3] Claudio Luchinat,[1,4] Giacomo Parigi,[1,4] Jing Yuan[1]

[1]Magnetic Resonance Center (CERM), University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy.

*e-mail: ivanobertini@cerm.unifi.it

[2]Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy.

[3]Present address: Department of Chemistry, University of Ioannina, 45110 Ioannina, Greece

[4]Department of Agricultural Biotechnology, University of Florence, via Maragliano 75-77, 50144 Florence, Italy

3.3.1 Solution structure and dynamics of S100A5 in the apo and Ca^{2+} -bound states

INTRODUCTION

Calcium(II) ions play a central role in cell signalling. Calcium(II) is toxic at elevated levels to cellular metabolism, and therefore, its influx and efflux in the cytosol must be controlled and kept at submicromolar resting levels.[1] Members of the EF-hand family of proteins can bind calcium ions in highly conserved helix-loop-helix motifs (called EF-hand). EF-hand proteins are constituted by several EF-hand motifs connected by a linker. Canonical EF-hand proteins have calcium binding loops constituted by 12 residues; S100 proteins are a subgroup where the N-terminal EF-hand loop is constituted by 14 residues.[2-4]

Calcium(II) binding may induce protein conformational changes, that may correlate with binding of target enzymes involved in a wide variety of cellular processes. The helices in the EF-hand motifs may in fact have an almost antiparallel arrangement, called closed conformation, or an almost orthogonal arrangement, called open conformation.[5-7] The latter conformation exposes large hydrophobic clefts on the protein surface, which may act as binding region for a variety of targets. Proteins undergoing conformational changes upon calcium(II) binding in the helices orientation inside each EF-hand motifs are generally functionally related to activation of target proteins, whereas the function of proteins not undergoing conformational changes is that of calcium buffer and transport.[8;9]

In general, S100 proteins undergo relatively small structural changes upon calcium(II) binding.[10] They are constituted by two EF-hand motifs. The N-terminal EF-hand comprises helix I, the pseudo calcium-binding site I and helix II. It is separated by a flexible linker, called hinge loop, from the C-terminal EF-hand which comprises helix III, the calcium-binding site II and helix IV.

Calbindin D_{9k} is a 8.5 kDa protein (76 residues), belonging to the S100 subgroup.[2;4] It is the only monomeric S100 protein, and thus able to bind two calcium ions. Its structure is known both in the apo and in the dicalcium form.[11;12] These structures have revealed that there are very small changes in the conformation of the first EF-hand motif upon calcium(II) binding, and somewhat larger but still minor changes in the second EF-hand motif.[13] This can be ascribed to the fact that each EF-hand motif has a conformation that is already optimised

for calcium binding even in the apo form.

In contrast to the small structural changes accompanying calcium(II) binding, flexibility of calbindin was shown to be largely reduced upon calcium(II) binding;[14] in the loop of the second EF-hand motif, in particular, S^2 passes from 0.63 to 0.85. This supports the idea that internal motions are important factors in the calcium buffering function of calbindin.[14] The first EF-hand loop has substantially smaller fluctuations on the picosecond to nanosecond time scale than the second EF-hand loop, and the effect of ion binding on these fluctuations is smaller. This can be ascribed to the difference between a consensus-type EF-hand motif and the S100 N-terminal EF-hand motif.

Members of the S100 protein family have been found to be implicated in the Ca^{2+}-dependent (and, in some cases, Zn^{2+}- or Cu^{2+}-dependent) regulation of a variety of intracellular activities, and several biological target have been identified for the different proteins.[4] A large variability in the sequence is observed among the protein subgroup, which is causing modulation of the shape of the binding surface in order to bind different targets. As a consequence, the several high-resolution structures available for S100-target peptide adducts display a remarkable lack of uniformity in the orientation of the target.[10] Furthermore, individual S100 proteins can bind different targets in different ways.[15]

All the S100 structures determined to date, with the exception of calbindin D_{9k} show that these proteins exist as homodimer, heterodimers or tetramers.[2] Most of the S100 proteins are homodimer. The dimer interface consists of helices I (I') and IV (IV') of each monomer arranged in a X-type four-helix bundle, both in the apo and in the calcium-bound states.[2] Calcium binding results in only minor alterations of the backbone conformation of calcium-binding site I, as shown for calbindin D_{9k}, but causes helix III to reorient and form a more open structure with respect to the apo state. As a result, in several S100 proteins hydrophobic residues of helices III and IV in the C-terminal EF-hand are more exposed, thus facilitating the interaction with a variety of target proteins.

Most of the S100 proteins, among which S100A5, are encoded by genes located in the same chromosome 1q21, with the exception of genes encoding S100B (located on chromosome 21q22), calbindin D9k (located on chromosome Xp22) and S100P (located on chromosome 4p16).[16] Interestingly, S100A1, S100A3, S1004, S100A5, S100A6, S100A8, S100A9, S100A12 and S100A13 genes are all mapped within a short distance.

We report here the solution structure determination and mobility studies of the protein S100A5 in both the apo and the calcium(II)-bound states, and compare it to the other S100 proteins. S100A5 is a 10.8 kDa protein for monomer (92 residues), for which biological targets have not yet been identified.[3]

MATERIALS & METHODS

Sample preparation

Untagged human *S100A5* was cloned into the *Nde*I and *Bam*HI sites of the expression vector pET21a. The recombinant plasmid was transformed into *E. Coli* BL21 Gold cells. For the production of ^{15}N-labelled or ^{13}C and ^{15}N-labelled S100A5, cultures were grown in minimal medium using ^{15}N ammonium sulphate and/or ^{13}C glucose as the sole nitrogen and carbon source, respectively. Cells were grown at 37°C to an OD 600 of 0.6 and induced with 1 mM IPTG. After induction, the temperature was reduced to 25°C and the culture grown overnight. Cells were harvested by centrifugation at 15,000 *g* for 15 minutes and resuspended in lysis buffer (20 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 mM PMSF, 2 mM DTT. Cell lysis was performed by sonicating with 8 bursts of 30 seconds each. The suspension was ultracentrifuged at 200,000 *g*. for 30 minutes. The cleared lysate was precipitated by slowly adding ammonium sulphate to 30% and centrifuging at 15,000 *g* for 20 minutes. The supernatant was brought to 2 mM $CaCl_2$, applied to a phenyl sepharose column equilibrated with 20 mM Tris-HCl, pH 7.5, 2 mM $CaCl_2$. The unbound proteins were washed out from the column with the same buffer. S100A5 was then eluted with 20 mM Tris-HCl, pH 7.5, 5 mM EDTA. The elute was concentrated and applied to a Superdex 75 column equilibrated with 30 mM MES, pH 6.5, 100 mM NaCl, 5 mM DTT. The fractions containing S100A5 were pooled and washed with excess of EDTA to remove all metal ions. The yield of S100A5 was 20 mg per litre of culture.

NMR spectroscopy and structure determination

All NMR experiments for assignments were performed at 298 K on a Bruker 500 MHz spectrometer equipped with cryo-probe. Apo and Ca_2-S100A5 samples were ^{13}C, ^{15}N labeled, 0.4 mM in MES 30 mM, NaCl 100 mM and DTT 5 mM buffer, pH 6.5, containing 10% D_2O. Sequential assignments of the backbone resonance were achieved via HNCO, HNCA, CBCA(CO)NH and HNCACB spectra. Side chain assignments were performed through 3D (h)CCH-TOCSY, HBHA(CBCACO)HN together with ^{13}C-NOESY-HSQC and ^{15}N-NOESY

HSQC. Proton-proton distance restraints were derived from the analysis of 2D-NOESY, ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC acquired on a Bruker 900 MHz spectrometer equipped with cryo-probe. The spectra were processed using the TOPSPIN program and analyzed through CARA program.[17] The secondary structure elements were predicted from the chemical shift index and the backbone dihedral angles were obtained from TALOS,[18] accordingly. The structures were calculated using the program CYANA-2.0.[19] The two monomers in the dimeric structure were linked together through a chain of dummy atoms with zero van der Waals' radius. The calcium(II) ions were included in the calculation of the calcium-loaded form by adding new residues in the amino acid sequence. Four chains of dummy atoms with zero van der Waals' radius, that can freely penetrate into the protein, each of them ending with one atom with a radius of 1.8 Å, which mimics the calcium ion, were added. Protein ligand atoms were linked to the metal ion through upper distance limits of 3 Å, according to the structure of S100A13.

The best 30 structures out of the calculated 350 structure of the CYANA family were then subjected to restrained energy minimization (REM) with AMBER 10.[20] NOE and torsion angle constraints were applied with force constants of 50 kcal mol^{-1} Å^{-2} and 32 kcal mol^{-1} rad^{-2}, respectively. The program PROCHECK-NMR[21] was used to evaluate the quality of the structures.

Relaxation measurements

^{15}N-R_1, R_2, and steady-state heteronuclear ^{1}H–^{15}N NOEs were measured at on a 700 MHz spectrometer using standard pulse sequences,[22;23] at 298 K. The longitudinal (R_1) and transverse (R_2) relaxation rates were determined by fitting the cross-peak intensities as a function of the delay to a single-exponential decay through the standard routines of the Sparky software[24]. The heteronuclear NOE values were obtained from the ratio of the peak height for ^{1}H-saturated and unsaturated spectra. The heteronuclear NOE values and their errors were estimated by calculating the mean ratio and the standard error from the available data sets. Reliable R_1, R_2, and NOE values were obtained for 67 and 71 out of the 92 assigned backbone ^{N}H resonances for the apo and the calcium forms, respectively. Estimates of the reorientational time were then calculated with the model-free approach.[25] Theoretical predictions of ^{N}H R_1 and R_2 values for apoS100A5 and Ca_2-S100A5 were calculated by using the HYDRONMR software.[26]

Metal binding detection

Binding of apo-S100A5 to Ca^{2+} was detected by following the chemical shift of protein NMR peaks in ^{1}H-^{15}N HSQC spectra during the titration with each metal ion. The chemical shift perturbation corresponding to each residue between the free and bound states was monitored by calculating the composite chemical shifts according to Eq. (1)

$$\delta(HN) = \sqrt{\frac{\Delta\delta_H{}^2 + (\Delta\delta_N / 5)^2}{2}} \qquad (1)$$

where $\Delta\delta_H$ and $\Delta\delta_N$ are the differences in the chemical shift of the amide proton and of the nitrogen, respectively, observed between the bound and free states.

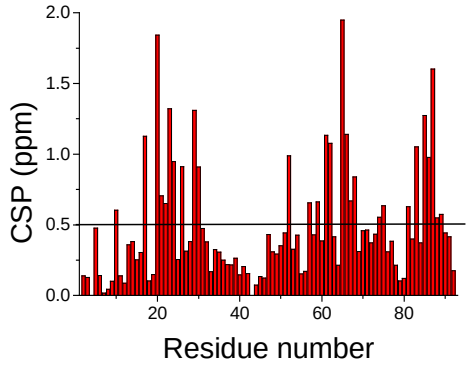
RESULTS

Resonance assignment

The ^{1}H ^{15}N HSQC of S100A5 in both the apo and the calcium forms show well dispersed resonances, as expected for a regularly folded protein. The backbone resonance signals were assigned from residue Glu 2 to residue Tyr 83 and from residue Phe 87 to residue Lys 92 in the apo form, and from residue Glu 2 to residue Lys 92 in the calcium form. The NMR spectra used for the assignments are indicated in the Materials and Methods section.

Ca^{2+} titration of apo-S100A5

The binding of apo-S100A5 to calcium(II) was monitored by following the changes in the ^{1}H-^{15}N HSQC NMR spectrum of ^{15}N-labelled apoS100A5 upon addition of increasing amounts of calcium(II). New peaks appeared in the spectrum during the titration corresponding to the calcium(II) bound S100A5 form. The intensity of the new peaks increased with increasing the Ca^{2+}:apo-S100A5 ratio. When the 2:1 ratio (with respect to the protein monomer concentration) was reached, the original peaks, corresponding to the apo form, disappeared. Such behaviour is indicative of a slow exchange regime, i.e. the exchange rate between the metal-free and the metal-bound forms is much smaller than the chemical shift difference between two forms. Figure 1 shows the chemical shift perturbation between the apo and calcium forms of S100A5. The residues with the chemical shifts mostly different in the two forms are located in the Ca^{2+} binding loops of the two EF-hand motifs, as expected, and also in the C-terminus. However, significant differences occur throughout the whole protein, thus indicating a significant conformational change in the protein structure in the calcium form with respect to the apo form.



Figue 1: Composite chemical shift perturbation, as defined in Eq. 1, of apoS100A5 upon calcium(II) binding. The horizontal line indicates the average value.

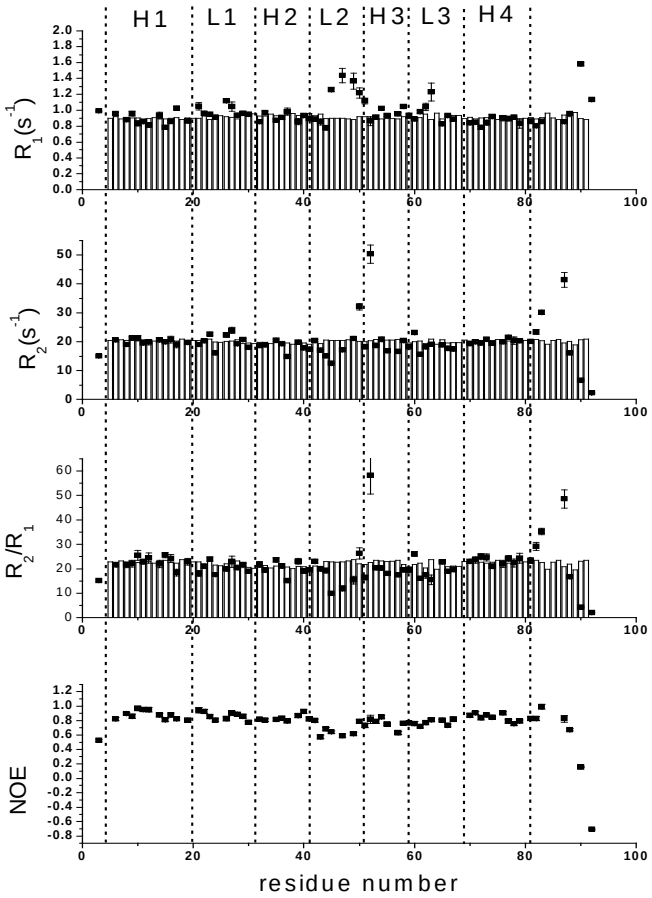
Relaxation measurements

The relaxation parameters for apo and the calcium-loaded S100A5 are shown in Figure 2. Such measurements indicate that the protein is dimeric in both forms. The reorientational times corresponding to the observed relaxation rates were in fact calculated to be 12.6±1.0 and 13.5±1.8 ns for the apo and calcium-loaded forms, respectively.

In both apo-S100A5 and Ca_2-S100A5, the relaxation rate measurements show large mobility in a time scale shorter than the reorientational time (R_1 increases, R_2 decreases, NOE decreases) in the hinge loop and for the last residues at the C-terminal, thus indicating that such regions may be largely unstructured. Some residues of the calcium-binding loops (21, 26, 27, 61-63 in the apo form; 25, 27, 30 in the calcium form) also show occurrence of motion.

In the apo protein, motion in a slower time scale (μs-ms) was observed for some residues localized at the beginning of helix III (Asp50 and Ile52), at the end of helix IV (Tyr83), and at the C-terminal (Phe87), as indicated by the significantly larger R_2 value (for Ile52, in particular R_2 is 52 s^{-1} vs. an average value of about 20 s^{-1} observed for helix residues). This slow motion may originate from backbone amide conformational exchange and/or side chain rotation. Since the time scale for the conformational exchange (μs-ms) is largely shorter than the mixing time (100 ms) of NOESY spectra, for these residues some ^1H-^1H NOEs might be observed between the side chains and residues localized in different directions upon conformational reorientations (see later). Upon calcium binding, no residues in slow motion are observed any more. However, calcium binding does not reduce the fast motion detected for residues in the hinge loop, the observed ^1H-^{15}N NOE being even smaller than in the apo form (the average ^1H-^{15}N NOE in the hinge loop is 0.60 and 0.40 for the apo and calcium form, respectively).

A) ApoS100A5



B) Ca_2S100A5

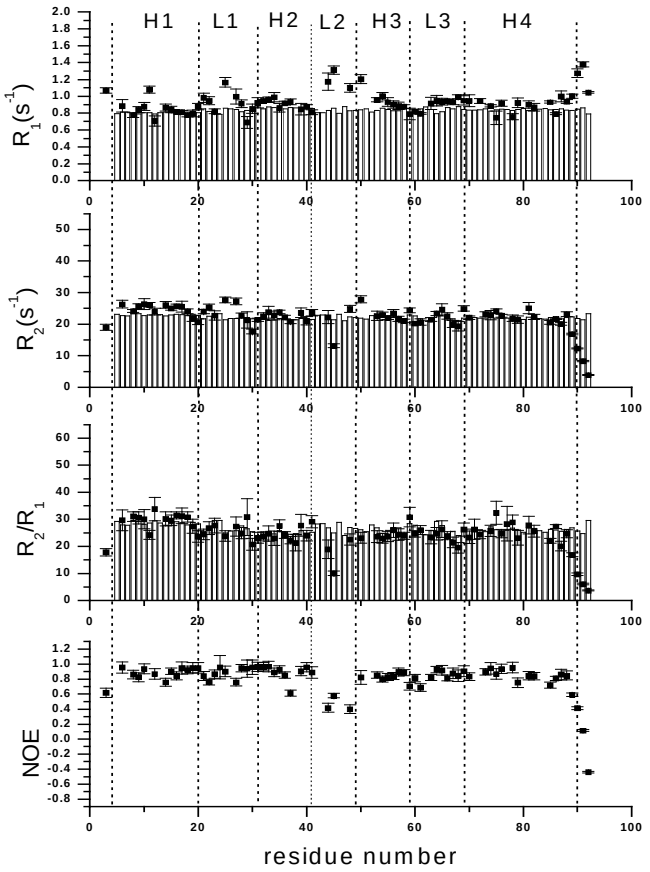


Figure 2 Sequential plot of relaxation parameters of apo and Ca^{2+}-S100A5

Solution structures of apo and Ca^{2+}-bound S100A5

The solution structures of the human S100A5 in the apo and calcium-loaded forms were obtained. A total of 2752 and 2530 meaningful upper distance limits for dimer, including 184 and 190 inter-monomer upper distance limints, for the apo and the calcium form were used, respectively (Tables 1 and 2). Few NOE patterns were detected for residues in the hinge loop and at the C-terminal, consistently with the observed mobility of such regions.

The calculated families of structures are shown in Figure 3. In both forms, the eight individual helices of the 2 EF-hand motifs of each monomer forming the dimeric structure are very well defined; the four Ca-binding loops are less well defined, whereas the linker regions between the two EF-hand Ca-binding domains are poorly defined. These results are in line with what previously found for other EF-hand proteins.[27;28] Each calcium-binding loop contains a short antiparallel β strand.

The root mean square deviation (RMSD) to the mean monomeric structure for the structured region of the protein is 0.73 ± 0.10 Å for the backbone and 1.22 ± 0.09 Å for all heavy atoms for apo-S100A5 (residues 3-40, 53-82) and 0.71 ± 0.09 Å for the backbone and 1.20 ± 0.07 Å for all heavy atoms for Ca_2-S100A5 (residues 3-40, 50-90). The RMSD to the mean dimeric structure for the structured region of the protein is 0.78 ± 0.09 Å for the backbone and 1.25 ± 0.09 Å for all heavy atoms for apo-S100A5 (residues 3-40, 53-82 of both monomers) and 0.83 ± 0.10 Å for the backbone and 1.29 ± 0.09 Å for all heavy atoms for Ca_2-S100A5 (residues 3-40, 50-90 of both monomers). More than 95% of residues (including those in the poorly defined regions) in all structures were located in the allowed regions of the Ramachandran plot. The conformational and energetic analysis of both structures are reported in Tables 1 and 2.

The relaxation rates were then calculated using HydroNMR[26] and the calculated structures, and are reported in Figure 2 as bars. An overall agreement is observed between calculated and experimental values for the residues located on the protein helices; on the other hand, differences between calculated and observed values make easier to appreciate the presence of mobility in some residues of the loops.

In both the apo and calcium-loaded forms, S100A5 forms homodimers due to the interactions between helices I and IV of each monomer. The symmetric relationship between the monomers is formed by a twofold rotational axis passing through the dimer interface

approximately perpendicular helix I and helix I' and parallel to helix IV and helix IV'. At the dimer interface, residues in the hinge loop between helix II and helix III make contacts with residues near the N terminus of helix I of the other monomer. The residues Phe 69, Lys 70, Ser73 and Cys 80 in helix IV also make several contacts with helix I' and helix IV' of the other monomer. All these interaction align helix I and helix IV in opposite directions to helix I' and helix IV', respectively, in the dimer.

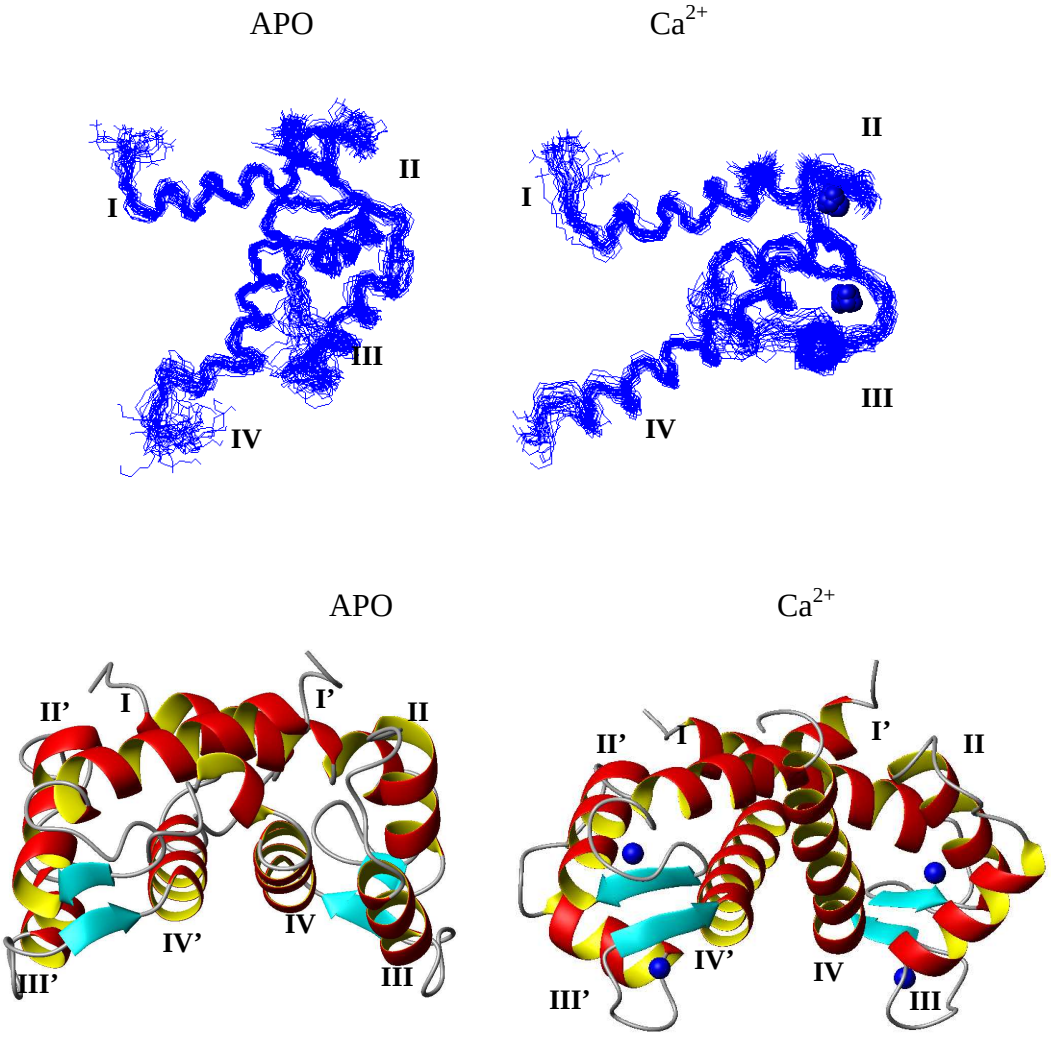


Figure 3. The solution structures of S100A5

DISCUSSION

The overall structures of both the apo and the calcium(II)-loaded forms of S100A5 are in good agreement with those obtained for other S100 proteins, like S100A1, S100A4, S100A6, S100A8, S100A12, S100A13 or S100B.[29-31] Comparison of the apo and calcium-loaded S100A5 structures shows that the N-terminal EF-hands (residues 5-41) are similar among

them (the backbone RMSD is 2.0 Å), thus indicating that there is no large conformational rearrangement upon calcium binding. In contrast, the C-terminal EF-hand (residues 49-82) undergoes a major conformational change upon calcium binding, the backbone RMSD between the two forms increasing to 4.2 Å. This conformational rearrangement includes a quite different orientation of helix III and somewhat changes in helix IV and in the hinge loop (Figure 4). These rearrangements upon calcium binding are essentially the same as those observed for other S100 proteins,[29;30;32-34] with the exception of S100A10 and Calbindin D_{9k} which are known to have a 'calcium-ready state' both in the N-terminal and the C-terminal EF-hands even in the absence of calcium(II) coordination. In apo-S100A6 and apo-S100A13, for instance, helix III is almost antiparallel to helix IV, but opens to 30-40° upon calcium binding. The same degree of opening is observed in other EF-hand proteins, like calmodulin,[6] not belonging to the S100 family. In S100A5 the angles between helix III and IV change from 168° to 118° on passing from the apo to the calcium-bound form, so that the two helices are close to be perpendicular in the latter form.

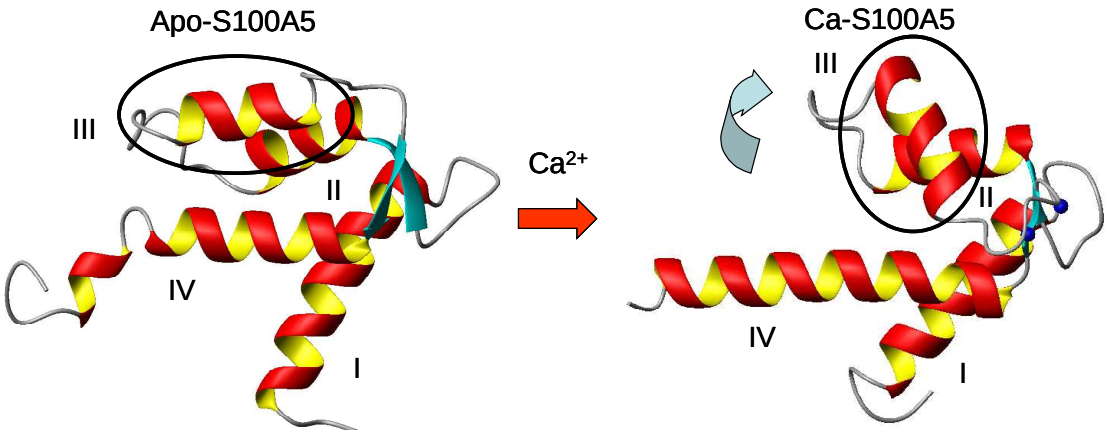


Figure 4. Major structural differences upon calcium binding are the different angle between helix III and IV and the longer alpha helical structure of helix IV.

Analogously to most S100 proteins, helices IV and IV' in the apo form tend to be antiparallel (forming an angle of 152°), whereas they form an angle of about 130° in the calcium-bound form, while helices I and I' form a similar angle (147°-142°) in both forms.

Structural changes within the EF-hand family can be monitored through a principal component analysis of the six interhelix angles representing the reciprocal orientation of the four helices.[6] It was shown that the EF-hand proteins can be clustered according to subgroups and metal content using the first two principal components, which concentrate the information distributed throughout the six interhelix angles. The values of the first two principal

components also permit to identify whether S100 proteins have a structure typical of apo or calcium-loaded form. The principal component values have thus been calculated for the obtained two forms of S100A5, and plotted together with the values previously calculated for the other S100 proteins (Figure 5), and with the values relative to other S100 proteins deposited in PDB, by using the same coefficients for the interhelix angles reported in [6]. The figure shows that S100A5 is regularly positioned regarding to the other S100 proteins both in the apo and in the calcium-loaded forms, thus pointing out to the occurrence of similar structures, and thus of similar rearrangement upon calcium(II) binding. To be noted that the only two S100 proteins not regularly placed are Calbindin D_{9k} and S100A10 in the apo form.

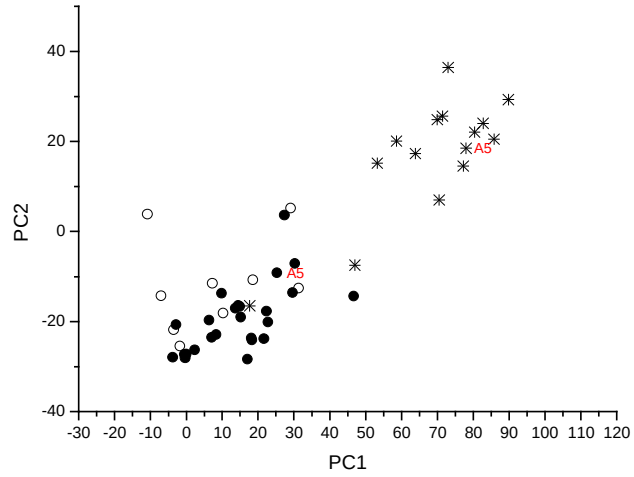


Figure 5. Principal component plot for the S100 proteins derived from principal components analysis of the six interhelix angles.[6] Apo proteins are indicated with an asterix, calcium-loaded proteins with a solid symbol and peptide-bound S100 proteins with an open symbol.

The concomitant 50° reorientation of helix III with respect to helix IV and the reorientation and translation of helices IV and IV' in S100A5 upon calcium(II) binding results in an increased solvent exposed surface of the hinge loop and of some positively charged residues of helix III and helices II in the calcium-loaded form. In fact, several hydrophobic residues on helix III (Ile 52, Leu 55, Met 56, and Leu 59), helix II (Ile 38) and helix IV (Phe 75, Met 78, Tyr 83) are constrained in a hydrophobic cluster in the apo-S100A5, which is loosen upon calcium(II) binding. On the other hand, calcium binding results in a decrease in exposure of metal ligand residues Asp60, Asn 62, Asp64, and Glu71 in the C-terminal calcium binding loop. The structural differences induced by calcium(II) binding in the homodimer thus lead to an exposure of two symmetrically positioned clefts defined by helix III, helix IV, the hinge loop and the last C-terminal residues, similarly to what found for other S100 proteins, where target proteins can be accommodated.[2]

In the apo form, residue Ile52 was identified by relaxation measurements to experience mobility on the millisecond to microsecond time scale. The side chains of these residue experience NOE contacting with both Met56 and Tyr83, which are positioned in opposite directions. As anticipated in the previous section, this may be due to occurrence of conformational exchange, and these data thus indicate that the side chain of Ile52 can rotate along an axis perpendicular to helix III, so that a conformational exchange is also affecting the backbone amide group. As a consequence, the residues forming a hydrophobic patch with Ile52, and particularly the ones localized on the protein surface like residue Tyr83, may also experience sizable mobility on side chain and/or backbone atoms.

The slow time scale motion detected for residue Ile52 and Tyr83 in the apo form is not revealed any more in the calcium form, as a consequence of conformational rearrangement of both helix III and helix IV. In the apo form, in fact, the aromatic ring of Tyr 83 of each monomer participates in a hydrophobic cluster including the side chains of residues Leu 44, Met 47, Ile 52 and Leu 79 of the same monomer. As already noted,[29] calcium binding overcomes the hydrophobic interactions that keep this cluster together, so that the side chain of Tyr 83 changes orientation and forms new hydrophobic contacts with the side chains of Leu 9, Val13 and Thr14 of helix I of the other monomer of the dimeric structure. This rotation is experimentally confirmed by the NOE observed between tyrosine 83 and residues Leu 44, Lys 48, Ile 52 and Leu 79 of the same monomer in the apo form and with residues Leu 9, Thr 10 and Val 13 of the other monomer in the calcium(II) form. The loosening of the hydrophobic cluster including Leu 44 and Met 47 may be responsible for the larger fast motion observed for the hinge loop residues from the lower NOE values measured for the calcium form with respect to those of the apo form.

Another difference between apo-S100A5 and Ca_2-S100A5 is that the C-terminal helix IV is shorter in the apo form. This very same difference has already been observed for S100A6 and S100B.[28;29;35] This is due to the unwinding of helix IV in apo-S100A5 at Tyr 83. The different orientation of the side chain of tyrosine 83 is in fact responsible for a break in the alpha helical structure, being consistent with a regularly formed alpha helix only in the calcium-loaded form.[29]

The hydrophobic residues at the extreme C-terminus (Phe 87 and Leu 88) are important for stabilizing both the apo- and the Ca_2-S100A5 homodimer. However, interestingly, these

residues form a hydrophobic cluster with different partners. In the apo form, they are in contact with Leu 27 in the first calcium binding loop of the other monomer, whereas in the calcium form they have hydrophobic interactions with Val 13 and Thr 14 of helix I of the other monomer.

The combination of these structural differences results in a change in the global shape and charge distribution of the S100A5 homodimer (see Figure 6). The change in the shape of the protein on passing from the apo to the calcium-loaded form is common to most of S100 proteins.[2;30] The change in the charge distribution depends largely on the particular S100 protein.[36] S100A5 shows a large charge distribution on the protein surface both in the apo and in the calcium-loaded form. Such electrostatic surface is somewhat smaller than in the apo form of S100A1, S100A4 and S100A13, but sizably larger than in the apo for of S100A2 and S100A6. Interestingly, upon calcium binding, some more exposed positive residues (Lys 48, Lys 57) are moved away from the opened cleft, which becomes slightly more hydrophobic, and the negative electrostatic surface is more clustered around the calcium binding sites. The total hydrophobic surface on the protein dimer is however reduced in the calcium(II)-loaded form with a concomitant increase of the charged surfaces. Other S100 proteins show a different change in the surface charge and hydrophobic distribution upon calcium binding: for instance, in S100A4, a larger increase of the hydrophobic surface was observed; in S100B, a larger negative charged surface is exposed.

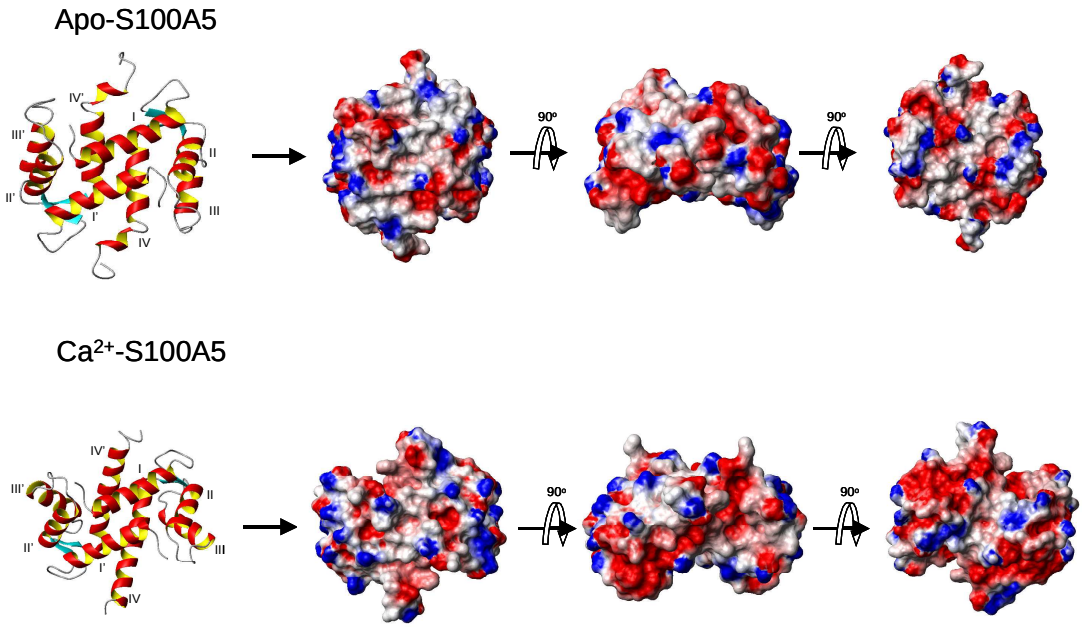


Figure 6. Electrostatic surface representation of the S100A5 dimers

Analogously to apoS100B and apoS100A4,[37;38] relaxation studies indicate that helix I and IV are quite rigid, whereas helix III is slightly more flexible. On the other side, in S100A5 the loop experiencing sizable mobility is only the hinge loop, whereas in S100A4 both the hinge loop and the calcium binding loops are quite mobile and in S100B the mobility of the calcium binding loops is even larger than that of the hinge loop. [37;38] A larger mobility for the hinge loop than for the calcium binding loops was also observed for S100A1.[27]

Titration experiments indicate that the affinity for calcium(II) is quite large for both the binding sites. This is in contrast with what observed for other S100 proteins, such as S100A13, where the calcium(II) affinity of binding site I is lower.[3;30]

Table 1. Statistical analysis of Apo-A100A5

	Family	***Mean***
The number of NOE upper distance limits		
Intra-monomer	1284	1284
Intra-residue	358	358
Inter-residue		
Sequential (|i-j| = 1)	395	395
Medium-range (|i-j| < 4)	320	320
Long-range (|i-j > 5)	211	211
Inter-monomer	184	184
Total dihedral angle restraints		
Phi	96	96
Psi	96	96
RMS violations per meaningful distance constraint (Å) :		
Intraresidue	0.0183± 0.0028	0.0178
Sequential	0.0179± 0.0025	0.0176
Medium range	0.0149± 0.0023	0.0093
Long range	0.0090± 0.0020	0.0087
RMS violations per meaningful dihedral angle constraints (°):		
Phi	4.87± 1.23	4.00
Psi	4.04± 1.42	2.82
Average number of constraints per residue	14.96	14.96
Average number of violations per conformer:		
Phi	6.93± 2.00	8.00
Psi	6.63± 2.20	4.00
NOE violations between 0.1 Å and 0.3 Å	9.40± 2.59	8.0
NOE violations larger than 0.3 Å	0	0
Average RMSD to the mean (Å)		
Backbone	1.00± 0.09[a] 0.78± 0.09[b]	
Heavy	1.47± 0.09[a] 1.25 ± 0.09[b]	
Residual CYANA Target Function ($Å^2$)	1.18 ± 0.31	
Structural analysis		
% of residues in most favorable regions	82.7 [a]/88.1 [b]	92.6
% of residues in allowed regions	13.2 [a]/10.0[b]	6.6
% of residues in generously allowed regions	2.2[a]/0.9[b]	0.0
% of residues in disallowed regions	1.9 [a]/1.0 [b]	0.8

a RMSD value were calculated in the sequence range 3-82.

b RMSD value were calculated exclude the flexible loop 41-52 of both two monomers.

Table 2. Statistical analysis of Ca_2-A100A5

	Family	***Mean***
The number of NOE upper distance limits		
Intra-monomer	1170	1170
Intra-residue	380	380
Inter-residue		
Sequential (|i-j| = 1)	325	325
Medium-range (|i-j| < 4)	280	280
Long-range (|i-j > 5)	185	185
Inter-monomer	190	190
Total dihedral angle restraints		
Phi	120	120
Psi	120	120
RMS violations per meaningful distance constraint (Å) :		
Intraresidue	0.0106± 0.0027	0.0123
Sequential	0.0092± 0.0016	0.0077
Medium range	0.0085± 0.0018	0.0083
Long range	0.0056± 0.0018	0.0053
RMS violations per meaningful dihedral angle constraints (°):		
Phi	1.97± 0.23	1.6246
Psi	0.63± 0.42	0.5218
Average number of constraints per residue	13.75	13.75
Average number of violations per conformer:		
Phi	9.74± 1.76	9.0
Psi	1.83± 1.23	2.0
NOE violations between 0.1 Å and 0.3 Å	4.03± 1.93	6.0
NOE violations larger than 0.3 Å	0	0
Average RMSD to the mean (Å)		
Backbone	0.93± 0.11[a] 0.83± 0.10[b]	
Heavy	1.40± 0.10[a] 1.29 ± 0.09[b]	
Residual CYANA Target Function ($Å^2$)	0.31 ± 0.03	
Structural analysis		
% of residues in most favorable regions	86.0 [a]/90.4 [b]	90.7
% of residues in allowed regions	11.2 [a]/7.5 [b]	9.3
% of residues in generously allowed regions	1.9 [a]/1.5 [b]	0.0
% of residues in disallowed regions	0.9 [a]/0.6 [b]	0.0

a RMSD value were calculated in the sequence range 3-90.

b RMSD value were calculated exclude the flexible loop 41-49 of both two monomers.

Reference List

(1) Orrenius, S.; Zhivotovsky, B.; Nicotera, P. *Nature Rev.Mol.Cell Biol.* **2003**, *4*, 552-565.

(2) Donato, R. *Int.J.Biochem.Cell Biol.* **2001**, *33*, 637-668.

(3) Santamaria-Kisiel, L.; Rintala-Dempsey, A. C.; Shaw, G. S. *Biochemical Journal* **2006**, *396*, 201-214.

(4) Donato, R. *Biochim.Biophys.Acta* **1999**, *1450*, 191-231.

(5) Ikura, M. *Trends Biochem.Sci.* **1996**, *21*, 14-17.

(6) Babini, E.; Bertini, I.; Capozzi, F.; Luchinat, C.; Quattrone, A.; Turano, M. *J.Proteome Res.* **2005**, *4*, 1961-1971.

(7) Capozzi, F.; Luchinat, C.; Micheletti, C.; Pontiggia, F. *J.Proteome Res.* **2007**, *6*, 4245-4255.

(8) Bertini, I.; Gupta, Y. K.; Luchinat, C.; Parigi, G.; Peana, M.; Sgheri, L.; Yuan, J. *J.Am.Chem.Soc.* **2007**, *129*, 12786-12794.

(9) Fragai, M.; Luchinat, C.; Parigi, G. *Acc.Chem.Res.* **2006**, *39*, 909-917.

(10) Bhattacharya, S.; Bunick, C. G.; Chazin, W. J. *Biochim.Biophys.Acta* **2004**, *1742*, 69-79.

(11) Bertini, I.; Donaire, A.; Jiménez, B.; Luchinat, C.; Parigi, G.; Piccioli, M.; Poggi, L. *J.Biomol.NMR* **2001**, *21*, 85-98.

(12) Svensson, L. A.; Thulin, E.; Forsén, S. *J.Mol.Biol.* **1992**, *223*, 601-606.

(13) Akke, M.; Forsén, S.; Chazin, W. J. *J.Mol.Biol.* **1995**, *252*, 102-121.

(14) Akke, M.; Skelton, N. J.; Kördel, J.; Palmer, A. G., III; Chazin, W. J. *Biochemistry* **1993**, *32*, 9832-9844.

(15) Bhattacharya, S.; Large, E.; Heizmann, C. W.; Hemmings, B.; Chazin, W. J. *Biochemistry* **2003**, *42*, 14416-14426.

(16) Ridinger, K.; Ilg, E. C.; Niggli, F. K.; Heizmann, C. W.; Schäfer, B. W. *Biochim.Biophys.Acta* **1998**, *1448*, 254-263.

(17) Keller, R. *The Computer Aided Resonance Assignment Tutorial*; CANTINA Verlag: Goldau, 2004.

(18) Cornilescu, G.; Delaglio, F.; Bax, A. *J Biomol NMR* **1999**, *13*, 289-302.

(19) Guntert, P. *Methods Mol.Biol.* **2004**, *278*, 353-378.

(20) Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M, Wang, B., Pearlman, D. A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C. E., Caldwell, J. W., Ross, W. S., and Kollman, P. A. AMBER 10. (8.0). 2008. San Francisco, CA, University of California.

Ref Type: Computer Program

(21) Laskowski, R. A.; Rullmann, J. A. C.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. *J.Biomol.NMR* **1996**, *8*, 477-486.

(22) Kay, L. E.; Torchia, D. A.; Bax, A. *Biochemistry* **1989**, *28* , 8972-8979.

(23) Barbato, G.; Ikura, M.; Kay, L. E.; Pastor, R. W.; Bax, A. *Biochemistry* **1992**, *31*, 5269-5278.

(24) Goddard, T. D. and Kneller, D. G. SPARKY 3, University of Califofnia, San Francisco. 2000.
Ref Type: Computer Program

(25) Lipari, G.; Szabo, A. *J.Am.Chem.Soc.* **1982**, *104*, 4546-4559.

(26) Garcia de la Torre, J. G.; Huertas, M. L.; Carrasco, B. *J.Magn.Reson.* **2000**, *147*, 138-146.

(27) Zhukov, I.; Ejchart, A.; Bierzynski, A. *Biochemistry* **2008**, *47*, 640-650.

(28) Smith, S. P.; Shaw, G. S. *Structure* **1998**, *6*, 211-222.

(29) Otterbein, L.; Kordowska, J.; Witte-Hoffmann, C.; Wang, C. L.; Dominguez, R. *Structure* **2002**, *10*, 557-567.

(30) Arnesano, F.; Banci, L.; Bertini, I.; Fantoni, A.; Tenori, L.; Viezzoli, M. S. *Angew.Chem.Int.Ed Engl.* **2005**, *44*, 6341-6344.

(31) Drohat, A. C.; Baldisseri, D. M.; Rustandi, R. R.; Weber, D. J. *Biochemistry* **1998**, *37*, 2729-2740.

(32) Marenholz, I.; Heizmann, C. W.; Fritz, G. *Biochem Biophys Res Commun* **2004**, *322*, 1111-1122.

(33) Maler, L.; Sastry, M.; Chazin, W. J. *J.Mol.Biol.* **2002**, *317*, 279-290.

(34) Bhattacharya, S.; Chazin, W. J. *Structure* **2003**, *11*, 738-739.

(35) Kilby, P. M.; Van Eldik, L. J.; Roberts, G. C. *Structure* **1996**, *4*, 1041-1052.

(36) Koch, M.; Diez, J.; Fritz, G. *J.Mol.Biol.* **2008**, *378*, 933-942.

(37) Inman, K. G.; Baldisseri, D. M.; Miller, K. E.; Weber, D. J. *Biochemistry* **2001**, *40*, 3439-3448.

(38) Dutta, K.; Cox, C. J.; Basavappa, R.; Pascal, S. M. *Biochemistry* **2008**, *47*, 7637-7647.

3.3.2 Implications on zinc binding to S100A5

Several S100 proteins bind not only Ca^{2+} but also Zn^{2+} with high affinity.[1] For some of them, such as S100B and S100A12, the calcium binding affinity is increased in the presence of zinc.[2,3] For some other, such as S100A2, zinc binding prevents calcium binding, thereby inhibiting the response to intra-cellular calcium signals.[4] S100A3 binds Ca^{2+} with low affinity (K_D = 4-35 mM), but binds four Zn^{2+} ions per dimer with exceptionally high affinity (K_D = 4 nM). The crystal structure actually shows a large distortion of the C-terminal canonical EF hand motif, which most likely makes Ca^{2+} binding difficult. The high affinity for zinc is attributed to the high cysteine content of S100A3, [5] which contain 10 cysteine residues. In this thesis, the binding mode for Zn^{2+} has also been performed for both apo and Ca_2-S100A5.

During the titration of Zn^{2+} to ^{15}N apoS100A5, the chemical shifts were observed to change during the titration, indicating a 1:1 binding in the Zn^{2+}:apo-S100A5 monomeric concentration. Several peaks of S100A5 shift with increasing Zn^{2+} concentration, being indicative a fast exchange on NMR time scale for Zn^{2+} binding. In this case, the combined chemical shift perturbations (CSP) are quite smaller (10 times less) than those observed when titrating with Ca^{2+} (Figure 7A), thus indicating that only small conformational changes occur in apo-S100A5 structure upon binding with Zn^{2+}. The CSP higher than the average (0.065 ppm) are clustered in the residue ranges 2-10, 20-26, 41-52, 68-70 and 80-87. For systems in fast exchange, the dissociation constant, K_D, was determined by fitting the CSP as a function of the metal:protein concentration.[6] From the analysis of the chemical shift perturbation, a dissociation constant of $(7.6\pm0.2)\times10^{-6}$ M is calculated using the same protocol used in ref 6 and the fitting curve is in Figure 8. As discussed in Section 3.3.1, S100A5 binds Ca^{2+} in a process slow on the NMR time scale. These data indicate that apoS100A5 has a higher affinity for Ca^{2+} than for Zn^{2+} ions. This estimation of affinity is similar to what found in the previous work.[5]

Increasing amounts of Zn^{2+} ions were also added stepwise to the Ca_2-S100A5 sample, and still zinc binding occurs. In this case, the most largely perturbed chemical shifts are located around cysteine 43 (Figure 7B). The line broadening of the peaks in the ^{1}H-^{15}N HSQC NMR spectrum however prevented us from further characterization.

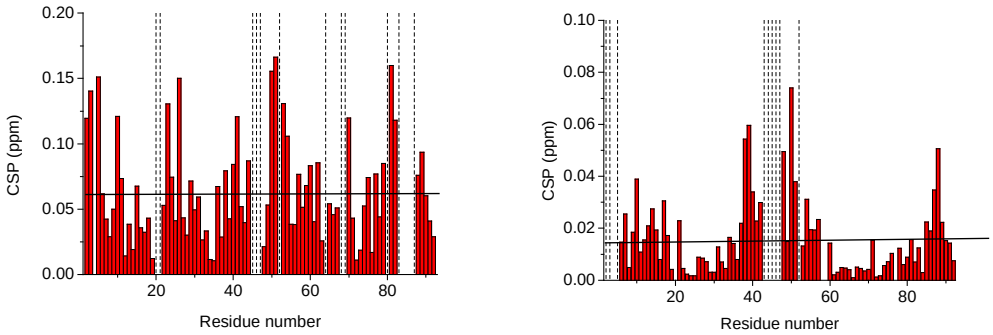


Figure 7: A) Composite chemical shift perturbation of apoS100A5 upon zinc(II) binding; B) composite chemical shift perturbation of Ca_2-S100A5 upon zinc(II) binding. The dashed bars indicate the residues disappeared in the Zn-bound protein. The horizontal lines indicate the average value.

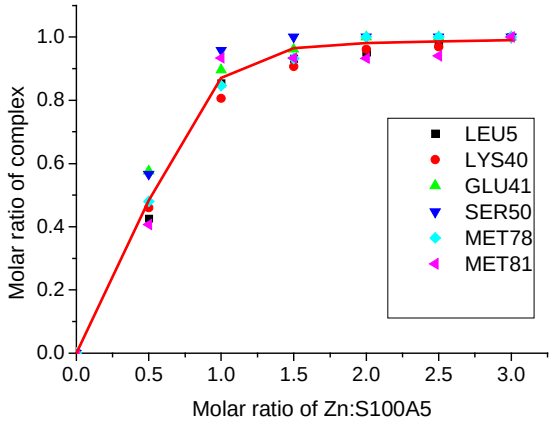


Figure 8. The chemical shift perturbation of the residues with the largest change upon zinc(II) binding as a function of the Zn:S100A5 concentration is fitted to Eq. (2) to obtain the dissociation constant.

Relaxation measurements performed for ZnS100A5 provide a reorientational times of 14.1±1.0 ns, thus pointing out that the protein maintains a homodimeric structure and sizable mobility in the hinge loop and the last residues at the C-terminal. (Figure 9) The line-width of several resides were too broad to detect, indicating that those residues are possibly sampling multiple conformations, exchanging with an intermediate rate.

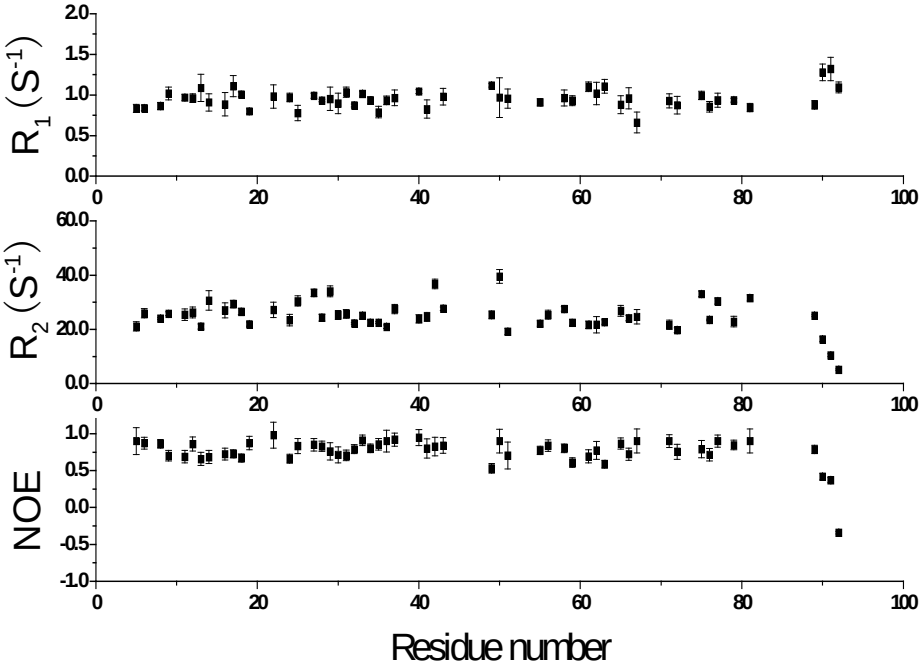
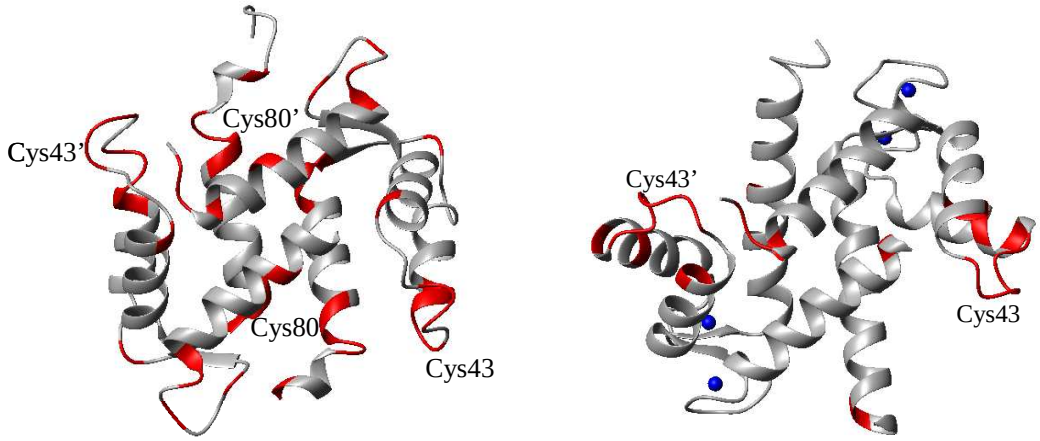


Figure 9: Sequential plot of relaxation parameters of ZnS100A5

Mapping the residues with the largest perturbation upon Zn^{2+} binding (Figure 10A) on the apo-S100A5 dimeric structure provides the indication that the possible Zn-binding site is located on the hinge loop and/or the last part of helix IV. We may hypothesize that both Cys 43 (in the hinge loop) and Cys 80 (in helix IV) are the ligands.



A) B)

Figure 10: Mapping the residues (red) with the largest perturbation upon zinc(II) binding on A) the apo-S100A5 dimeric structure and b) Ca_2-S100A5 dimeric structure

Ca_2S100A5 can also bind Zn^{2+}. The difference in the chemical shifts during titration with Zn^{2+} indicates that the latter ion is bound to Cys 43(Figure 10B). Such residue is located on

the surface of the protein, so that several dimers may be linked by a Zn^{2+} ion, and multimers can in this case be formed. The formation of multimers would be responsible for the degradation in the quality of the acquired spectra. In Ca_2S100A5, in fact, Cys 80 is located too far from Cys 43 due to the reorientation of helix IV with respect to the apo structure.

Reference list

(1) Heizmann, C. W., and Cox, J. A. *Biometals.* **1998,** 11, 383–397

(2) Baudier, J., Glasser, N., and Gerard, D. *J. Biol. Chem.* **1986**, 261, 8192–8203

(3) Dell'Angelica, E. C., Schleicher, C. H., and Santome, J. A. *J. Biol. Chem*.**1994**,269, 28929–28936

(4) Franz, C., Durussel, I., Cox, J. A., Schafer, B. W., and Heizmann, C. W. *J. Biol. Chem*. **1998**, 273, 18826–18834

(5) Fritz, G., Mittl, P.R., Vasak, M., Grutter, M.G., and Heizmann, C.W. *J. Biol. Chem*.**2002**, 277, 33092–33098

(6) Deep, S.D., Im, S.C., Zuiderweg, E.R.P., Waskell, L. *Biochemisty*, **2005**, 44, 10654-10668

4
CONCLUSIONS AND PERSPECTIVE

MAP program was developed to provide maximum allowed probabilities (MAPs) of conformations in protein domains not rigidly connected. The approach is applied to free CaM and to its adduct with α-synuclein. The largest MAP values were obtained for both systems, by using the pcs and rdc data measured from Tb^{3+}, Tm^{3+}, and Dy^{3+} of N60D-CaM. The MAP analysis for CaM provides further information on the variety of conformations experienced by the system. Such variety is somewhat reduced in the CaM-α-synuclein adduct, which however still retains high flexibility. The power of MAP approach 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 less ghosts. This results in higher accuracy in the identification of the conformational space experienced. This approach is particularly useful for proteins with domains experiencing flexibility, such as multidomain proteins, and for protein-protein adducts.

For the rigid system, a strategy was developed to improve the accuracy of a protein structure in solution is to take a crystal structure as a starting model and to “correct” it by applying pcs and rdc data. The structural characterization of CaM with two peptides representing the interaction surface of two protein partners, the DAPk and DRP-1 peptide were investigated. The spread of the rdc valus is the same for the two domains of CaM, which indicates the two domains, are blocked. However, analyses of rdc values indicate that the solution structures are similar, but not identical, to the crystal structures. Some discrepancies with respect to the crystal structure are pinpointed, that can be ascribed to a structural rearrangement from solid state to solution. Comparative examination of the crystal and solution structures suggests that the main origin for the discrepancy may lie in the loss of the intermolecular hydrogen bonds. The combined approach of crystal structure and analysis of paramagnetic restraints provides solution structures as precise as the crystallographic structure and quite accurate. The approach can also be applied to refine single domain proteins containing a metal ion binding site, either natural.

The solution structures of S100A5 were determined for both the apo and the Ca^{2+} form by NMR spectroscopy. Both two forms are homodimers in solution. The structural differences induced by Ca^{2+} binding in the homodimer thus lead to an exposure of two symmetrically

positioned clefts defined by helix III, helix IV, the hinge loop and the last C-terminal residues, similarly to what found for other S100 proteins, where target proteins can be accommodated. Dynamic properties are also different between the apo and Ca^{2+} form. Some residues were detected slow time scale motion on both helix and loop region, while fast motion is present in the hinge region for Ca^{2+} bound form.

In the future, the study of the complex of CaM binding with full-length CaM regulatory domain DAPk or DRP-1 kinase, together with their catalytic domain will be also attempted. Once the crystal/NMR structures of these complex will be determined in the active state, the intriguing mechanism of autoinhibition of this Ca^{2+}/CaM-refulated kniase will be deciphered at the atomic level. The specific structural features, which are responsible for this unique autoinhibitory mechanism, may provide an interesting common hallmark for DAPk or DRP-1 kinase which should distinguish them from the other CaM-regulated kinases. Another interesting issue relates to investigating multiple domain orientations of CaM when it binds full-length CaM regulatory domain DAPk or DRP-1 kinase by using MAP approach.

5
Source code of MAP program

```
      PROGRAM MAP
      implicit double precision(a-h,o-z)
      parameter(maxmaxfn=500,maxmaxii=10)
      dimension conf(maxmaxfn,7*maxmaxii)
      dimension nconf(maxmaxfn)
      dimension orient(maxmaxfn,4)
      integer npoints,ipoints,maxii
      common /points/ conf,orient,nconf,maxii,npoints,ipoints
      include 'annealingparam.f'
      integer lexpdir
      character*100 expfile,inline,expdir,infile,outfile,resfile
      common /files/ expfile,inline,expdir,infile,outfile,resfile,lexpdir
           common flag
      integer n,ndim
      dimension guess(3),x(3),xx(1000)
      integer exitcode
      integer maxit,maxfn
      integer errcode
      external threshpmax
      external fcnall

      errcode=0
      expfile="./expfile"
      open(unit=1,file=expfile,status='old',err=1000)
      errcode=1
      read(1,'(a)',err=1000,end=1000) inline
      errcode=2
      ii=1
      call extractstring(inline,expdir,ii,jj)
      if (jj.eq.0) goto 1000
      lexpdir=jj-1
      close(1)
      infile="./"//expdir(1:lexpdir)//"/"//"infile"
      outfile="./"//expdir(1:lexpdir)//"/"//"outfile"
      errcode=3
      open(unit=1,file=infile,status='old',err=1000)
      errcode=4
      read(1,*,err=1000,end=1000) guess(1),guess(2),guess(3)
      close(1)
      errcode=5
      npoints=0
      open(unit=1,file="./results0",status='old',err=10)
      ijk=1
      do
         read(1,*,end=10,err=10) (orient(ijk,i),i=1,4)
         read(1,*,end=10,err=10) nconf(ijk)
         do i=1,7*(nconf(ijk)-1)+3-1
            read(1,*,end=10,err=10) conf(ijk,i)
         end do

         ind=0
         do i=1,ijk-1
            temp=dabs(orient(i,1)-orient(ijk,1))+
     1         dabs(orient(i,2)-orient(ijk,2))+
     2         dabs(orient(i,3)-orient(ijk,3))
            if (temp.lt.0.1d0.or.orient(ijk,4).gt.0.3d0 ) then
               ind=1
               exit
            end if
         end do
         if (ind.eq.0) ijk=ijk+1
         npoints=ijk
         if (npoints.gt.maxmaxfn) goto 10
      end do
10    continue
      if (npoints.gt.0) npoints=npoints-1
      maxfn=200
```

```
            maxii=maxmaxii
            ipoints=0
            n=3
            ftol=1d-4
            maxit=30
             s=1.d0
            errcode=6
      open(unit=2,file=outfile,status='replace',err=1000)
            write(2,*) "Starting orient:",(guess(ijk),ijk=1,3)
            write(2,*) "Initial points read:",npoints

            call threshpmax(n,guess,fvalue)
!     call dsimplex(threshpmax,n,guess,s,ftol,maxfn,x,fvalue)

            print *,"init=",(guess(i),i=1,3)," final=",(x(i),i=1,3)
            print *,"fvalue=",fvalue," exitcode=",exitcode
            write(2,*) "Final orient:",(x(ijk),ijk=1,3)
            write(2,*) "Final TF:",fvalue
            write(2,*) "Final npoints:",npoints
            close(2)
            errcode=7
             resfile="./"//expdir(1:lexpdir)//"/"//"results05"
            open(unit=1,file=resfile,status='unknown',err=1000)
            do ijk=1,npoints
               if(ijk .eq.npoints)then
                  write(1,*) (orient(ijk,i),i=1,4)
                  write(1,*) nconf(ijk)
                  do i=1,7*(nconf(ijk)-1)+3-1
                       write(1,*) conf(ijk,i)
                   end do
               end if
          if (orient(ijk,4).gt.0.5d0) cycle
          ind=0
          do i=1,ijk-1
             if (dabs(orient(ijk,1)-orient(i,1))+dabs(orient(ijk,2)-orient(i,2))+
     1       dabs(orient(ijk,3)-orient(i,3)).lt.0.1d0) then
               ind=1
              exit
             end if
          end do
          if (ind.eq.1) cycle
               write(1,*) (orient(ijk,i),i=1,4)
               write(1,*) nconf(ijk)
               do i=1,7*(nconf(ijk)-1)+3-1
                  write(1,*) conf(ijk,i)
               end do
          end do
          close(1)
          stop
1000  continue
      if (errcode.eq.0) then
         write(*,*) "Error opening expfile:",expfile
      else if (errcode.eq.1) then
         write(*,*) "Error reading from expfile:",expfile
         write(*,*) "expdir=",expdir
      else if (errcode.eq.2) then
         write(*,*) "Error parsing:",expdir
      else if (errcode.eq.3) then
         write(*,*) "Error opening infile:",infile
      else if (errcode.eq.4) then
         write(*,*) "Error reading from infile:",infile
      else
         write(*,*) "Unidentified error: code=",errcode
      end if
           end
           subroutine threshpmax(nnn,v1,pvalue)
        INCLUDE 'sup6rdc.h'
```

```
      include 'annealingparam.f'
        dimension npt(MAXFE,6),
     *      ccptx(MAXSTR,MAXFE,6), ccpty(MAXSTR,MAXFE,6),
     *      ccptz(MAXSTR,MAXFE,6),
     *      vx(3), vy(3),vz(3),nresno(MAXFE)
        character namepara*4
        character yes*1
           parameter(maxmaxfn=500,maxmaxii=10)
           dimension nconf(maxmaxfn)
           dimension conf(maxmaxfn,7*maxmaxii)
           dimension orient(maxmaxfn,4)
           dimension Dx(100),Dy(100),Dz(100)
           integer ncheck
           integer npoints,ipoints,maxii
           common /points/ conf,orient,nconf,maxii,npoints,ipoints
           common m
          common/position/pos1,pos2,pos3,C4,wpcs,wrdc
          common /pp/ Dx,Dy,Dz
          common /number/ncheck
          common /metal/x2,y2,z2
           common/iprint/iprint
           dimension xxguess(1000),xx(1000),xxsave(1000),v1(3)
      real*8 temp
      integer rjj
       external fcnall
      logical foundAT(MAXFE,3)
        gltoll=0.
        field=700.
        costrdc=1./(5.892E-6*field**2/42.578**2)
        filename1='TbTmDyC.pcs'
        filename2='bestbaxCSDR.pdb'
        filename3='out1bax'
        filename4='TbTmDyC.rdc'
        filename5='out2bax'
        fileout='out'
        nat=2528          !numero atomi
        factorrdc=0.02
        nstr=1
        namepara='ME'
        PERC=0
           intsys=0
           nsystem=1
        ngrid=1000            !numeri calcoli
c            *****        READ OBSERVED SHIFT ******
        open (1,file=filename1,status='old')
        i=0
        numS=0
        numD=0
        numT=0
        print *,'        *****        READ OBSERVED SHIFT'
        do while (1.gt.0)
           read(1,'(a)',end=99) line
           i=i+1
      read(line,'(i4,1x,a3,2x,a4,f9.3,4x,i2,f6.2,f10.3)')
     *                numres(i),namres(i),namat(i),obs(i),
     *                mlprot(i),tolprot(i),wprot(i)
           if (tolprot(i).eq.0) tolprot(i)=gltoll
           if ((PERC.gt.0).or.(PERC.le.1)) then
              ttperc=abs(obs(i)*PERC)
              if (ttperc.gt.tolprot(i)) then
                    tolprot(i)=ttperc
              end if
           end if
           if (mlprot(i).eq.0.or.mlprot(i).eq.1) then
              mlprot(i)=1
              numS=numS+1
           else
```

```
            if (mlprot(i).eq.2) numD=numD+1
            if (mlprot(i).eq.3) numT=numT+1
          end if
          if (wprot(i).eq.0.) wprot(i)=1.0
          if (wprot(i).eq.-1.) wprot(i)=0.0
        enddo
 99     ihp=i
        print *, '         *****         TOTAL DIPOLAR PROTONS ',ihp
        open (2,file=filename2,status='old')
        open (3,file=filename3)
        nfe=0
        tmpnfe=0
c            *****         READ COORDINATE FILE                *****
        print *,'         *****         READ COORDINATE FILE'
        ncontpoint=0
        do 10 k=1,nstr
       do 20 i=1,nat
         read(2,'(a)',end=98) line
            read(line,'(5x,i6,1x,a4,6x,i4,4x,3f8.3)')
     *                        numberatom,nameatom,nores,xa,ya,za
          if ((nameatom(1:2).eq.namepara(1:2).or.
     *          nameatom(2:3).eq.namepara(1:2))) then
              tmpnfe=tmpnfe+1
          fx(k,tmpnfe)=xa
          fy(k,tmpnfe)=ya
          fz(k,tmpnfe)=za
        end if
              read(line,'(5x,i6,1x,a4,1x,a3,1x,i5,4x,3f8.3)')
     *             num_at(k,i),nam_at(k,i),nam_res(k,i),num_res(k,i),
     *             xp(k,i),yp(k,i),zp(k,i)
 20   continue
        nfe=tmpnfe
        tmpnfe=0
 10     continue
        print *, '         *****         TOTAL PARAMAGN. CENTERS ',nfe

        print *,'         *****         WRITE OBSERVED OUT FILE'
        do 140 k=1,nstr
          icontprot=0
          do 150 j=1,ihp
         do 160 i=1,nat
              if (num_res(k,i).eq.numres(j)) then
                if (nam_at(k,i).eq.namat(j))then
                  icontprot=icontprot+1
             write(3,
     *       '(i6,1x,a4,1x,i3,1x,a3,1x,3F8.3,1x,F7.3,1x,i1,2(1x,f7.3))')
     *             num_at(k,i),nam_at(k,i),num_res(k,i),nam_res(k,i),
     *           xp(k,i),yp(k,i),zp(k,i),obs(j),mlprot(j),
     *             tolprot(j),wprot(j)
            goto 150
           end if
                end if
 160       continue
           print *,namat(j),numres(j),namres(j)
 150 continue
 140    continue
 98     continue
        if (icontprot.lt.ihp) then
          print *,' ****** ERROR:    SOME PCS NOT FOUND '
          print *,' '
          print *,' ****** PROGRAM STOP'
          print *,' '
          STOP
        endif
        close(1)
        close(2)
        close(3)
```

```
      nhp=icontprot
c                     *****         READ OBSERVED DeltaJs ******
      open (1,file=filename4,status='old')
      i=0
      numS=0
      numD=0
      numT=0
      print *,'         *****         READ OBSERVED DeltaJs'
      do while (1.gt.0)
         read(1,'(A)',end=199) line
         i=i+1
       read(line,'(i4,1x,a3,2x,a4,f9.3,4x,i2,f6.2,f10.3)')
     *                numres(i),namres(i),namat(i),obs(i),
     *                mlprot(i),tolprot(i),wprot(i)
         if(namat(i)(1:1).eq.' ') then
            namat(i)(1:3)=namat(i)(2:4)
            namat(i)(4:4)=' '
         end if
         if (tolprot(i).eq.0) tolprot(i)=gltoll
         if ((PERC.gt.0).or.(PERC.le.1)) then
            ttperc=abs(obs(i)*PERC)
            if (ttperc.gt.tolprot(i)) then
                tolprot(i)=ttperc
            end if
         end if
         if (wprot(i).eq.0.) wprot(i)=1.0
         if (wprot(i).eq.-1.) wprot(i)=0.0
      enddo
  199   ihp=i
      print *, '         *****         TOTAL No. NHs ',ihp
      open (2,file=filename2,status='old')
      open (3,file=filename5)
      print *,'         *****         READ COORDINATE FILE'
      ncontpoint=0
      do 410 k=1,nstr
      do 420 i=1,nat
      read(2,'(a)',end=198) line
            read(line,'(5x,i6,1x,a4,1x,a3,1x,i5,4x,3f8.3)')
     *              num_at(k,i),nam_at(k,i),nam_res(k,i),num_res(k,i),
     *              xp(k,i),yp(k,i),zp(k,i)
  420    continue
  410   continue
      print *,'         *****         WRITE PROCESSED INPUT FILE'
      do 240 k=1,nstr
         icontprot=0
         icont15n=0
         do 250 j=1,ihp
                  do 260 i=1,nat
               if (num_res(k,i).eq.numres(j)) then
                  if (nam_at(k,i).eq.namat(j))then
                     icontprot=icontprot+1
                           write(3,
     *         '(i6,1x,a4,1x,i3,1x,a3,1x,3F8.3,1x,F7.3,1x,i1,2(1x,f7.3))')
     *              num_at(k,i),nam_at(k,i),num_res(k,i),nam_res(k,i),
     *              xp(k,i),yp(k,i),zp(k,i),obs(j),mlprot(j),
     *              tolprot(j),wprot(j)
                     do ii=1,nat
                        if (num_res(k,ii).eq.numres(j))then
                        if (nam_at(k,ii)(1:3).eq.' N '
     *                       .or.nam_at(k,ii)(1:2).eq.'N ') then
                          icont15n=icont15n+1
                          write(3,
     *         '(i6,1x,a4,1x,i3,1x,a3,1x,3F8.3,1x,F7.3,1x,i1,2(1x,f7.3))')
     *           num_at(k,ii),nam_at(k,ii),num_res(k,ii),nam_res(k,ii),
     *           xp(k,ii),yp(k,ii),zp(k,ii),obs(j),mlprot(j),
     *              tolprot(j),wprot(j)
                          goto 250
```

```
                        end if
                        end if
                    end do
                    print*, '1',nam_at(k,i)(2:3),numres(j),namres(j)
                    goto 250
                          end if
               end if
260            continue
           print*, '2',namat(j),numres(j),namres(j)
250      continue
240    continue
198    continue
       if (icontprot.lt.ihp) then
          print *,' ****** ERROR:    SOME ATOMS NOT FOUND '
          print *,' '
          print *,' ****** THE PROGRAM WILL STOP'
          print *,' '
          STOP
       endif
       if (icontprot.ne.icont15n) then
          print *,' ****** ERROR: No. OF 15N ATOMS NOT EQUAL',
     *    ' TO No. OF HN'
          print *,' '
          print *,' ****** THE PROGRAM WILL STOP'
          print *,' '
          print *,icontprot,icont15n
          STOP
       endif
       close(1)
       close(2)
       close(3)
       nhprdc=icontprot*2
       open (1,file=filename3,status='old')
       i=0
       do 170    k=1,nstr
          do 170    i=1,nhp
             read(1,'(A)',end=299) line
             read(line,'(20x,3F8.3,F8.3,i2,2(1x,f7.3))')
     *                    cx(k,i),cy(k,i),cz(k,i),
     *                    obs((k-1)*nhp+i),mlprot((k-1)*nhp+i),
     *                    tolprot((k-1)*nhp+i),wprot((k-1)*nhp+i)
 170   continue
 299   continue
       close(1)
       open (1,file=filename5,status='old')
       i=0
       do 180    k=1,nstr
          do 180    i=1,nhprdc
             read(1,'(A)',end=399) line
             read(line,'(20x,3F8.3,F8.3,i2,2(1x,f7.3))')
     *                  cxrdc(k,i),cyrdc(k,i),czrdc(k,i),
     *                  obsrdc((k-1)*nhprdc+i),mlprotrdc((k-1)*nhp+i),
     *                  tolprotrdc((k-1)*nhprdc+i),wprotrdc((k-1)*nhprdc+i)
 180   continue
 399   continue
       close(1)
       sumtmp1=0
       do 3 i=1,nhp
            sumtmp1=sumtmp1+obs(i)**2
3      continue
       wpcs=1d0/sumtmp1
       print *, wpcs
       sumtmp1=0
       do 4 i=1,nhprdc-1,2
            sumtmp1=sumtmp1+obsrdc(i)**2
4      continue
       wrdc=1d0/sumtmp1
```

```
          print *,wrdc

        flag=0
        fvalueold=9999999999.
                    i=1
                    n=0
                    open(1,file='checkpoint',status='old')
                       do
                         read(1,*,err=5,end=5)Dx(i),Dy(i),Dz(i)
                         i=i+1
                         n=n+1
                       end do
5               continue
                   close(1)
                     ncheck=n
                     x2=-26.875+21.743
                     y2=-4.660 -5.455
                     z2=-0.478 -1.307

ccccccc          begin to calculate ccccccccccccc
C maxkkk          - number of steps to reach    pmaxthresh
C ndim0           - number of conformations of    goodstart
C pmaxthresh - threshold of pmax
C tfthresh      - threshold of TF
C nstep           - max number of minimizations for each C4
           nnpoints=0
           maxkkk=1
          tresh1=1d-3
                pmaxthresh=0.05d0
                tfthresh=0.15d0
                dpmax=pmaxthresh/maxkkk
                nstep=2500
                C4=dpmax
                pos1=v1(1)
                pos2=v1(2)
                pos3=v1(3)
                print*,pos1,pos2,pos3
CCC reading from database
                if(npoints.eq.0) then
                    i=2
                    ndim=(i-1)*7-1+3
                    xxguess(1)=4d0+5d0*rand()
                    xxguess(2)=4d0+5d0*rand()
                    xxguess(3)=4d0+5d0*rand()
                    xxguess(4)=acos(1.-2.*rand())
                    xxguess(5)=asin(1.-2.*rand())
                    xxguess(6)=asin(1.-2.*rand())
                    xxguess(7)=4d0+5d0*rand()
                    xxguess(8)=4d0+5d0*rand()
                    xxguess(9)=4d0+5d0*rand()
                  else
                    kk=0
                    temp2=1d9
                    do ijk=1,npoints
                       ndim0=(nconf(ijk)-1)*7-1+3
                       do k=1,ndim0
                          xx(k)=conf(ijk,k)
                       end do

                       call fcnall(ndim0,xx,temp)
                       write(*,*) "Point:",ijk," TF=",temp
                       if (temp.lt.temp2) then
                          temp2=temp
                          kk=ijk
                       end if
                    end do
                    if (kk.gt.0) then
```

```
                    i=nconf(kk)
                    ndim=(i-1)*7-1+3
                    do k=1,ndim
                       xxguess(k)=conf(kk,k)
                    end do
                    write(*,*) "Best start is point ",kk,":",(
     1                 orient(kk,ijk),ijk=1,3)
                    write(*,*) "Starting TF=",temp2
                  end if
                end if
ccccc       try to guess good translation ccccc
                  do ijk=1,10000
                    do kk=1,3
                       xx(kk)=xxguess(kk)
                       xxguess(kk)=xxguess(kk)+1d-1*(rand()-0.5d0)
                    end do
                    call fcnall(ndim,xxguess,temp1)
                    if (temp1.lt.temp2) then
                       temp2=temp1
                    else
                       do kk=1,3
                          xxguess(kk)=xx(kk)
                       end do
                    end if
                 end do
                   call fcnall(ndim,xxguess,temp2)
                   print *,"start1=",temp2
                   flag=0
                  s=1d0
                  ftol=1d-6
                  maxfn=10000
               expected      =1d-10
               tempestart =10.d0
               deltamulti =0.9d0
               delta2multi=0.9d0
               maxraise      =1.05d0
               tempelast    =0.05d0
               deltalast    =0.5d0
               tempemin      =1.d-4
               tempeup       =0.05d0
               deltaup       =0.5d0
               maxsec = 1000
               nok = 40000/dsqrt(ipoints+1d0)
               nno = 40000/dsqrt(ipoints+1d0)
               nmx = 70000/dsqrt(ipoints+1d0)
                ipoints=ipoints+1
               write(*,*) "npoints=",npoints," nok=",nok
               indprint=0
               ncode=1
              fvalue=temp2
                   kkk=0
                   do
                      kkk=kkk+1
                      if (kkk.gt.maxkkk) then
                    pvalue=fvalue
                    exit
                  end if
                  C4old=C4
                       ndim=(i-1)*7-1+3
                       C4=dpmax*kkk
                       write(*,*)C4
                       istep=0

                       if (kkk.gt.1) then
                    do kk=2,i-1
                      xxguess(10+7*(kk-2))=xxguess(10+7*(kk-2))*(1d0-C4)/(1d0-C4old)
                    end do
```

```
          end if
          ind=0
          do
         tempestart=10./(2*i)
        tempelast=10./(1*i)
        tempemin=10./(10*i)
            ncode=1
            ndim=(i-1)*7-1+3
                 call fcnall(ndim,xxguess,temp1)
            write(*,*) "i=",i," start=",temp1
            oldfvalue=fvalue
            call annealing(fcnall,ndim,xxguess,xx,fvalue,ncode)
            do ijk=1,ndim
               xxguess(ijk)=xx(ijk)
            enddo
            flag=0
            do ijk=1,ndim
               xxsave(ijk)=xxguess(ijk)
            end do
            ijk=1
            rjj=0
            temp=0d0
            do
               j=3+7*ijk
               if (j.gt.ndim) exit
               temp1=dabs(xxguess(j))
               xxguess(j)=temp1
               if (temp1.lt.tresh1) then
                  do ii=j+1,ndim
                     xxguess(ii-7)=xxguess(ii)
                  end do
                  ndim=ndim-7
                  call fcnall(ndim,xxguess,temp2)
                  if (2d0*(temp2-fvalue).gt.
     *                 0.01d0*(dabs(fvalue)+dabs(temp2))) then
                     print *,"Conformation ",ijk+1+rjj,"    not"
     *                ," removed ",(temp2-fvalue)*100d0
                     do ii=j-6,ndim
                        xxguess(ii)=xxsave(ii)
                     end do
                     ijk=ijk+1
                     ndim=ndim+7
                  else
                     print *,"Conformation ",ijk+1+rjj,
     *                    " removed ",(temp2-fvalue)*100d0
                     do ii=j+1,ndim+7
                        xxsave(ii-7)=xxsave(ii)
                     end do
                     fvalue=temp2
                     rjj=rjj+1
                  end if
               else
                  temp=temp+temp1
                  ijk=ijk+1
               end if
            end do
            temp1=1d0-temp-C4
            if (temp1.lt.tresh1) then
               ndim=ndim-7
               call fcnall(ndim,xxguess,temp2)
               if (2d0*(temp2-fvalue).gt.
     *              0.01d0*(dabs(fvalue)+dabs(temp2))) then
                  print *,"Conformation ",ijk+1+rjj,"    not"
     *                 ," removed ",(temp2-fvalue)*100d0
                  ndim=ndim+7
               else
                  print *,"Conformation ",ijk+1+rjj,
```

```
*                     " removed ",(temp2-fvalue)*100d0
                      fvalue=temp2
                      rjj=rjj+1
                    end if
                  end if
                  print *,rjj," conformation removed"
                  i=i-rjj
                  call fcnall(ndim,xxguess,temp1)
                  print*,temp1
                  if (fvalue.lt.tfthresh .and.kkk.ne.maxkkk) then
                    write(*,*) "Value under the threshold"
                    write(*,*) "Jumping to next weight"
                    exit
                  end if
                  istep=istep+1
                  if (istep.le.nstep) then
                if (rjj.gt.0) then
                      cycle
                    else if (i.eq.maxii) then
                      if (dabs(oldfvalue-fvalue).lt.1d-3*dabs(fvalue)) then
                        pvalue=fvalue
                        goto 999
                      else
                        cycle
                      end if
                  end if
                    else if (rjj.gt.0 .or. i.eq.maxii) then
                      if (kkk.eq.maxkkk) then
                        pvalue=fvalue
                        goto 999
                      else
                        kkk=maxkkk-1
                        exit
                      end if
                    end if
                    tempweight=xxguess(10)
                k=2
                temp=xxguess(10)
                do kk=3,i-1
                  if (xxguess(10+7*(kk-2)).gt.temp) then
                    temp=xxguess(10+7*(kk-2))
                    k=kk
                  end if
                  tempweight=tempweight+xxguess(10+7*(kk-2))
                end do
                temp1=1d0-tempweight-C4
                if (temp1.gt.temp) then
                  temp=temp1
                  k=i
                end if
                    print *,"k=",k," i=",i
                    i=i+1
                    ndim=ndim+7
                xxguess(ndim-6)=temp1
                do ijk=1,6
                  xxguess(ndim-6+ijk)=xxguess(3+7*(k-2)+ijk)
                end do
                xxguess(7*(k-2)+10)=xxguess(7*(k-2)+10)/2d0
                  end do
9               format(f9.3,8(2x,f9.3))
              end do
999            continue
                write(*,97)pvalue,pos1,pos2,pos3,xx(1),xx(2),xx(3)
97            format(f9.3,7(2x,f9.3))
              if (npoints.lt.maxmaxfn) then
                npoints=npoints+1
                nnpoints=npoints
```

```
  else if (nnpoints.lt.maxmaxfn) then
     nnpoints=nnpoints+1
  else
     nnpoints=1
  end if
        do ijk=1,ndim
           conf(nnpoints,ijk)=xxguess(ijk)
        end do
     print *, "nnpoints=",nnpoints," i=",i,
&      " ndim=",ndim, " pvalue=",pvalue
        nconf(nnpoints)=i
        orient(nnpoints,1)=pos1
        orient(nnpoints,2)=pos2
        orient(nnpoints,3)=pos3
        orient(nnpoints,4)=pvalue
              flag=1
              call fcnall(ndim,xxguess,fvalue)
              flag=0
        return
  end

     subroutine fcnall(ndim,xx,fvalue)
  include 'sup6rdc.h'
  dimension vett(NP)
       dimension Dx(100),Dy(100),Dz(100)
     integer ncheck
     common m
         common/position/pos1,pos2,pos3,C4,wpcs,wrdc
         common /pp/ Dx,Dy,Dz
          common /number/ncheck
          common /metal/x2,y2,z2
         common/iprint/ iprint
     dimension xx(ndim)
     dimension P2(100),O2(100),T2(100)
     dimension fx2(100),fy2(100),fz2(100)
     dimension axx(100),axy(100),axz(100),ayx(100),ayy(100)
     dimension ayz(100),azx(100),azy(100),azz(100)
     dimension xapp(100),yapp(100),zapp(100)
     dimension scalx(100),scaly(100),scalz(100)
     dimension r(100),g1(100),g2(100),w(100)

  IVIOLATION=0
  TMP1=0
  i=1
  do while (i.le.nhp*nstr)
     shift(i)=0.0
     i=i+1
  enddo
  fvalue=0.d0
  do 1 n=1,nstr
  do 2 m=1,nfe
     P1=pos1
     T1=pos2
     O1=pos3
     fx1=xx(1)
     fy1=xx(2)
     fz1=xx(3)
      a1d0=9825.          !metallo 1 Tb
      a2d0=3692.
      P0=0.511
      T0=-0.47
      O0=-0.61
     a1d02=6767.          !metallo 2 Tm
     a2d02=2417.
     P02=-2.377
     T02=-1.170
     O02=0.697
```

```
      a1d03=8958.              !metallo Dy
      a2d03=-3974.
      P03=-3.209
      T03=0.210
      O03=-0.123
      maxres=80
     axx0=cos(P0)*cos(O0)
     axy0=sin(P0)*cos(O0)
     axz0=sin(O0)
     ayx0=-cos(T0)*sin(P0)-sin(O0)*cos(P0)*sin(T0)
     ayy0=cos(T0)*cos(P0)-sin(O0)*sin(P0)*sin(T0)
     ayz0=sin(T0)*cos(O0)
     azx0=sin(T0)*sin(P0)-sin(O0)*cos(P0)*cos(T0)
     azy0=-sin(T0)*cos(P0)-sin(O0)*sin(P0)*cos(T0)
     azz0=cos(T0)*cos(O0)
     axx02=cos(P02)*cos(O02)
     axy02=sin(P02)*cos(O02)
     axz02=sin(O02)
     ayx02=-cos(T02)*sin(P02)-sin(O02)*cos(P02)*sin(T02)
     ayy02=cos(T02)*cos(P02)-sin(O02)*sin(P02)*sin(T02)
     ayz02=sin(T02)*cos(O02)
     azx02=sin(T02)*sin(P02)-sin(O02)*cos(P02)*cos(T02)
     azy02=-sin(T02)*cos(P02)-sin(O02)*sin(P02)*cos(T02)
     azz02=cos(T02)*cos(O02)
     axx03=cos(P03)*cos(O03)
     axy03=sin(P03)*cos(O03)
     axz03=sin(O03)
     ayx03=-cos(T03)*sin(P03)-sin(O03)*cos(P03)*sin(T03)
     ayy03=cos(T03)*cos(P03)-sin(O03)*sin(P03)*sin(T03)
     ayz03=sin(T03)*cos(O03)
     azx03=sin(T03)*sin(P03)-sin(O03)*cos(P03)*cos(T03)
     azy03=-sin(T03)*cos(P03)-sin(O03)*sin(P03)*cos(T03)
     azz03=cos(T03)*cos(O03)

     axx1=cos(P1)*cos(O1)
     axy1=sin(P1)*cos(O1)
     axz1=sin(O1)
     ayx1=-cos(T1)*sin(P1)-sin(O1)*cos(P1)*sin(T1)
     ayy1=cos(T1)*cos(P1)-sin(O1)*sin(P1)*sin(T1)
     ayz1=sin(T1)*cos(O1)
     azx1=sin(T1)*sin(P1)-sin(O1)*cos(P1)*cos(T1)
     azy1=-sin(T1)*cos(P1)-sin(O1)*sin(P1)*cos(T1)
     azz1=cos(T1)*cos(O1)

     kk=4
     tempweight=0.
     do ii=1,(ndim+1-3)/7
     P2(ii)=xx(kk)
     T2(ii)=xx(kk+1)
     O2(ii)=xx(kk+2)
     fx2(ii)=xx(kk+3)
     fy2(ii)=xx(kk+4)
     fz2(ii)=xx(kk+5)
     if(ii.eq.(ndim+1-3)/7)then
     w(ii)=abs(1.-tempweight-C4)
     else
     w(ii)=abs(xx(kk+6))
     tempweight=tempweight+w(ii)
     endif
     kk=kk+7
     axx(ii)=cos(P2(ii))*cos(O2(ii))
     axy(ii)=sin(P2(ii))*cos(O2(ii))
     axz(ii)=sin(O2(ii))
     ayx(ii)=-cos(T2(ii))*sin(P2(ii))-
&      sin(O2(ii))*cos(P2(ii))*sin(T2(ii))
     ayy(ii)=cos(T2(ii))*cos(P2(ii))-
&       sin(O2(ii))*sin(P2(ii))*sin(T2(ii))
```

```
            ayz(ii)=sin(T2(ii))*cos(O2(ii))
            azx(ii)=sin(T2(ii))*sin(P2(ii))-
     &        sin(O2(ii))*cos(P2(ii))*cos(T2(ii))
            azy(ii)=-sin(T2(ii))*cos(P2(ii))-
     &        sin(O2(ii))*sin(P2(ii))*cos(T2(ii))
            azz(ii)=cos(T2(ii))*cos(O2(ii))
            enddo
            if(tempweight+C4.gt.1.)then
            fvalue=1E6*(tempweight-1.)**2
            if(fvalue.le.10000000)fvalue=10000000
            goto 999
            endif
ccccccccccccc Check good or not ccccccccccccccccc
              j=0
              do ijk=1,ncheck
                   do ii=1,(ndim+1-3)/7
                      xapp2=DX(ijk)-fx2(ii)-fx(n,m)
                      yapp2=DY(ijk)-fy2(ii)-fy(n,m)
                      zapp2=DZ(ijk)-fz2(ii)-fz(n,m)
                   r1=sqrt(xapp2**2+yapp2**2+zapp2**2)
            scalx2=(xapp2*axx(ii)+yapp2*axy(ii)+zapp2*axz(ii))
            scaly2=(xapp2*ayx(ii)+yapp2*ayy(ii)+zapp2*ayz(ii))
            scalz2=(xapp2*azx(ii)+yapp2*azy(ii)+zapp2*azz(ii))
              r2=sqrt((scalz2-z2)**2+(scalx2-x2)**2+(scaly2-y2)**2)
                    if(r1.lt.12.d0) j=1
                    if(r2.lt.12.d0) j=2
                    if ((r1.lt.12.d0).and.(r2.lt.12.0)) j=3
                    if(j.eq.1) fvalue=fvalue+(r1-12.d0)**2*100000.
                    if(j.eq.2) fvalue=fvalue+(r2-12.d0)**2*100000.
                    if(j.eq.3)
     &              fvalue=fvalue+(r1-12.d0)**2+(r2-12.d0)**2*100000.
                         if(flag.eq.1.and.j.ne.0)then
                             print*,r1,r2,ii+1,ijk,j
                             print*,DX(ijk),fx2(ii),x2
                             print*,DY(ijk),fy2(ii),y2
                             print*,DZ(ijk),fz2(ii),z2
                          endif
                 end do
               end do
ccccccccccccccccc finish to check ccccccccccccccccccccccccccccc
888         continue
            tmp2=0.0
            i=1
            ihp=(n-1)*nhp
            do 10 while (i.le.nhp)
              if(numres(ihp+i).le.maxres)then
              else
                   xapp1=CX(n,I)-fx(n,m)-fx1
                   yapp1=CY(n,I)-fy(n,m)-fy1
                   zapp1=CZ(n,I)-fz(n,m)-fz1
                   scalz1=(xapp1*azx1+yapp1*azy1+zapp1*azz1)
                   scalx1=(xapp1*axx1+yapp1*axy1+zapp1*axz1)
                   scaly1=(xapp1*ayx1+yapp1*ayy1+zapp1*ayz1)

                   do ii=1,(ndim+1-3)/7
                   xapp(ii)=CX(n,I)-fx(n,m)-fx2(ii)
                   yapp(ii)=CY(n,I)-fy(n,m)-fy2(ii)
                   zapp(ii)=CZ(n,I)-fz(n,m)-fz2(ii)
            scalz(ii)=(xapp(ii)*azx(ii)+yapp(ii)*azy(ii)+zapp(ii)*azz(ii))
            scalx(ii)=(xapp(ii)*axx(ii)+yapp(ii)*axy(ii)+zapp(ii)*axz(ii))
            scaly(ii)=(xapp(ii)*ayx(ii)+yapp(ii)*ayy(ii)+zapp(ii)*ayz(ii))
                   enddo
          if(mlprot(ihp+i).eq.1)then
                   zapp1=(scalx1*azx0+scaly1*azy0+scalz1*azz0)
                   xapp1=(scalx1*axx0+scaly1*axy0+scalz1*axz0)
                   yapp1=(scalx1*ayx0+scaly1*ayy0+scalz1*ayz0)
```

```
                do ii=1,(ndim+1-3)/7
                zapp(ii)=(scalx(ii)*azx0+scaly(ii)*azy0+scalz(ii)*azz0)
                xapp(ii)=(scalx(ii)*axx0+scaly(ii)*axy0+scalz(ii)*axz0)
                yapp(ii)=(scalx(ii)*ayx0+scaly(ii)*ayy0+scalz(ii)*ayz0)
                enddo
      a1d=a1d0
      a2d=a2d0
      endif
      if(mlprot(ihp+i).eq.2)then
                zapp1=(scalx1*azx02+scaly1*azy02+scalz1*azz02)
                xapp1=(scalx1*axx02+scaly1*axy02+scalz1*axz02)
                yapp1=(scalx1*ayx02+scaly1*ayy02+scalz1*ayz02)
                do ii=1,(ndim+1-3)/7
                zapp(ii)=(scalx(ii)*azx02+scaly(ii)*azy02+scalz(ii)*azz02)
                xapp(ii)=(scalx(ii)*axx02+scaly(ii)*axy02+scalz(ii)*axz02)
                yapp(ii)=(scalx(ii)*ayx02+scaly(ii)*ayy02+scalz(ii)*ayz02)
                enddo
      a1d=a1d02
      a2d=a2d02
      endif
      if(mlprot(ihp+i).eq.3)then
                zapp1=(scalx1*azx03+scaly1*azy03+scalz1*azz03)
                xapp1=(scalx1*axx03+scaly1*axy03+scalz1*axz03)
                yapp1=(scalx1*ayx03+scaly1*ayy03+scalz1*ayz03)
                do ii=1,(ndim+1-3)/7
                zapp(ii)=(scalx(ii)*azx03+scaly(ii)*azy03+scalz(ii)*azz03)
                xapp(ii)=(scalx(ii)*axx03+scaly(ii)*axy03+scalz(ii)*axz03)
                yapp(ii)=(scalx(ii)*ayx03+scaly(ii)*ayy03+scalz(ii)*ayz03)
                enddo
      a1d=a1d03
      a2d=a2d03
      endif
                r1=sqrt(xapp1**2+yapp1**2+zapp1**2)
                g11=(sqrt(3.)*zapp1-r1)*(sqrt(3.)*zapp1+r1)
                g21=(xapp1-yapp1)*(xapp1+yapp1)
                primo=(A1D*G11+1.5*A2D*G21)/r1**5
                sumshift=0.
                do ii=1,(ndim+1-3)/7
                r(ii)=sqrt(xapp(ii)**2+yapp(ii)**2+zapp(ii)**2)
                g1(ii)=(sqrt(3.)*zapp(ii)-r(ii))*(sqrt(3.)*zapp(ii)+r(ii))
                g2(ii)=(xapp(ii)-yapp(ii))*(xapp(ii)+yapp(ii))
                secondo=(A1D*G1(ii)+1.5*A2D*G2(ii))/r(ii)**5
                terzo=w(ii)*secondo
                sumshift=sumshift+terzo
                enddo
             shift(ihp+i)=shift(ihp+i)+C4*primo+sumshift
             endif
             I=I+1
 10      continue
 2     continue
!         if(flag.eq.1)write(*,*)'PCS'
      do 3 i=1,nhp
         tmp2=abs(shift(ihp+i)-obs(ihp+i))
     *                            -tolprot(ihp+i)
         if (tmp2.gt.0.0) then
            IVIOLATION=IVIOLATION+1
            TMP1=tmp1+tmp2**2*wprot(ihp+i)
         endif
!         if(flag.eq.1)write(*,*)shift(ihp+i),obs(ihp+i),
!      * dabs(obs(ihp+1))/dabs(shift(ihp+i))
 3     continue
 1     continue
         if(flag.eq.1)write(*,*)'TF PCS=',tmp1, tmp1*wpcs
      TMP3=0
      i=1
      do while (i.le.nhprdc*nstr)
         shift(i)=0.0
```

```
      i=i+1
    enddo
    do 111 n=1,nstr
      tmp4=0.0
      i=1
      ihp=(n-1)*nhprdc
      do 100 while (i.le.nhprdc)
               xapp1=CXrdc(n,I)-fx(n,1)-fx1
               yapp1=CYrdc(n,I)-fy(n,1)-fy1
               zapp1=CZrdc(n,I)-fz(n,1)-fz1
               scalz1H=(xapp1*azx1+yapp1*azy1+zapp1*azz1)
               scalx1H=(xapp1*axx1+yapp1*axy1+zapp1*axz1)
               scaly1H=(xapp1*ayx1+yapp1*ayy1+zapp1*ayz1)
               xapp1=CXrdc(n,I+1)-fx(n,1)-fx1
               yapp1=CYrdc(n,I+1)-fy(n,1)-fy1
               zapp1=CZrdc(n,I+1)-fz(n,1)-fz1
               scalz1N=(xapp1*azx1+yapp1*azy1+zapp1*azz1)
               scalx1N=(xapp1*axx1+yapp1*axy1+zapp1*axz1)
               scaly1N=(xapp1*ayx1+yapp1*ayy1+zapp1*ayz1)
               scalx1=scalx1H-scalx1N
               scaly1=scaly1H-scaly1N
          scalz1=scalz1H-scalz1N
          r1=sqrt(scalx1**2+scaly1**2+scalz1**2)
               if(r1.gt.1.08.or.r1.lt.0.94) then
               print*, '***** ERROR: r(H-N) has a wrong value: ',r1
               print*, '***** Check data relative to constraint ',
*              (i+1)/2
               print*, '***** The program will stop'
               STOP 1
             end if
               do ii=1,(ndim+1-3)/7
               xapp(ii)=CXrdc(n,I)-fx(n,1)-fx2(ii)
               yapp(ii)=CYrdc(n,I)-fy(n,1)-fy2(ii)
               zapp(ii)=CZrdc(n,I)-fz(n,1)-fz2(ii)
          scalz2H=(xapp(ii)*azx(ii)+yapp(ii)*azy(ii)+zapp(ii)*azz(ii))
          scalx2H=(xapp(ii)*axx(ii)+yapp(ii)*axy(ii)+zapp(ii)*axz(ii))
          scaly2H=(xapp(ii)*ayx(ii)+yapp(ii)*ayy(ii)+zapp(ii)*ayz(ii))
               xapp(ii)=CXrdc(n,I+1)-fx(n,1)-fx2(ii)
               yapp(ii)=CYrdc(n,I+1)-fy(n,1)-fy2(ii)
               zapp(ii)=CZrdc(n,I+1)-fz(n,1)-fz2(ii)
          scalz2N=(xapp(ii)*azx(ii)+yapp(ii)*azy(ii)+zapp(ii)*azz(ii))
          scalx2N=(xapp(ii)*axx(ii)+yapp(ii)*axy(ii)+zapp(ii)*axz(ii))
          scaly2N=(xapp(ii)*ayx(ii)+yapp(ii)*ayy(ii)+zapp(ii)*ayz(ii))
        scalx(ii)=scalx2H-scalx2N
        scaly(ii)=scaly2H-scaly2N
        scalz(ii)=scalz2H-scalz2N
            r(ii)=sqrt(scalx(ii)**2+scaly(ii)**2+scalz(ii)**2)
            if(r(ii).gt.1.08.or.r1.lt.0.94) then
               print*, '***** ERROR: r(H-N) has a wrong value: ',r(ii),ii
               print*, '***** Check data relative to constraint ', (i+1)/2
               print*, '***** The program will stop'
               STOP 1
            end if
               enddo
if(mlprotrdc(ihp+i).eq.1)then
               zapp1=(scalx1*azx0+scaly1*azy0+scalz1*azz0)
               xapp1=(scalx1*axx0+scaly1*axy0+scalz1*axz0)
               yapp1=(scalx1*ayx0+scaly1*ayy0+scalz1*ayz0)
               do ii=1,(ndim+1-3)/7
               zapp(ii)=(scalx(ii)*azx0+scaly(ii)*azy0+scalz(ii)*azz0)
               xapp(ii)=(scalx(ii)*axx0+scaly(ii)*axy0+scalz(ii)*axz0)
               yapp(ii)=(scalx(ii)*ayx0+scaly(ii)*ayy0+scalz(ii)*ayz0)
               enddo
a1d=a1d0/costrdc
a2d=a2d0/costrdc
endif
if(mlprotrdc(ihp+i).eq.2)then
```

```
                zapp1=(scalx1*azx02+scaly1*azy02+scalz1*azz02)
                xapp1=(scalx1*axx02+scaly1*axy02+scalz1*axz02)
                yapp1=(scalx1*ayx02+scaly1*ayy02+scalz1*ayz02)
                do ii=1,(ndim+1-3)/7
                zapp(ii)=(scalx(ii)*azx02+scaly(ii)*azy02+scalz(ii)*azz02)
                xapp(ii)=(scalx(ii)*axx02+scaly(ii)*axy02+scalz(ii)*axz02)
                yapp(ii)=(scalx(ii)*ayx02+scaly(ii)*ayy02+scalz(ii)*ayz02)
                enddo
      a1d=a1d02/costrdc
      a2d=a2d02/costrdc
      endif
      if(mlprotrdc(ihp+i).eq.3)then
                zapp1=(scalx1*azx03+scaly1*azy03+scalz1*azz03)
                xapp1=(scalx1*axx03+scaly1*axy03+scalz1*axz03)
                yapp1=(scalx1*ayx03+scaly1*ayy03+scalz1*ayz03)
                do ii=1,(ndim+1-3)/7
          zapp(ii)=(scalx(ii)*azx03+scaly(ii)*azy03+scalz(ii)*azz03)
          xapp(ii)=(scalx(ii)*axx03+scaly(ii)*axy03+scalz(ii)*axz03)
          yapp(ii)=(scalx(ii)*ayx03+scaly(ii)*ayy03+scalz(ii)*ayz03)
                enddo
      a1d=a1d03/costrdc
      a2d=a2d03/costrdc
      endif
                g11=(sqrt(3.)*zapp1-r1)*(sqrt(3.)*zapp1+r1)
                g21=(xapp1-yapp1)*(xapp1+yapp1)
                primo=(A1D*G11+1.5*A2D*G21)/r1**5
                sumshift=0.
                do ii=1,(ndim+1-3)/7
                    g1(ii)=(sqrt(3.)*zapp(ii)-r(ii))*(sqrt(3.)*zapp(ii)+r(ii))
                    g2(ii)=(xapp(ii)-yapp(ii))*(xapp(ii)+yapp(ii))
                    secondo=(A1D*G1(ii)+1.5*A2D*G2(ii))/r(ii)**5
                    terzo=w(ii)*secondo
                    sumshift=sumshift+terzo
                enddo
              shift(ihp+i)=shift(ihp+i)+C4*primo+sumshift
              i=i+2
 100       continue
!           if(flag.eq.1)write(*,*)'RDC'
        do 30 i=1,nhprdc-1,2
            tmp4=abs(shift(ihp+i)-obsrdc(ihp+i))-tolprotrdc(ihp+i)
            if (tmp4.gt.0.0) then
              IVIOLATION=IVIOLATION+1
              TMP3=tmp3+tmp4**2*wprotrdc(ihp+i)
            endif
!            if(flag.eq.1)write(*,*)shift(ihp+i),obsrdc(ihp+i)
 30      continue
 111      continue
          if(flag.eq.1)write(*,*)'TF RDC=',tmp3,tmp3*wrdc
777      continue
         fvalue=fvalue+wpcs*tmp1+wrdc*tmp3

!VINCOLO DISTANZA    CA    !anche plane dipende da pdb
                    x78=-2.206
                    y78=1.071
                    z78=-3.587
                    x81=2.743
                    y81=0.334
                    z81=-5.010
                xv78=x78-fx(1,1)
                yv78=y78-fy(1,1)
                zv78=z78-fz(1,1)
                xv81=x81-fx(1,1)-fx1
                yv81=y81-fy(1,1)-fy1
                zv81=z81-fz(1,1)-fz1
                scalz21=(xv81*azx1+yv81*azy1+zv81*azz1)
                scalx21=(xv81*axx1+yv81*axy1+zv81*axz1)
                scaly21=(xv81*ayx1+yv81*ayy1+zv81*ayz1)
```

```
      r1=sqrt((xv78-scalx21)**2+(yv78-scaly21)**2+(zv78-scalz21)**2)
         dist=9 !3*3
         distmin=4 !5.5
         plane0=-3.-fx(1,1) !l'elica non puo' tornare indietro:
         if(r1.gt.dist)then
      fvalue=fvalue+10*(r1-dist)**2
         endif
         if(r1.lt.distmin)then
      fvalue=fvalue+10*(r1-distmin)**2
         endif
                  do ii=1,(ndim+1-3)/7
                  xv81=x81-fx(1,1)-fx2(ii)
                  yv81=y81-fy(1,1)-fy2(ii)
                  zv81=z81-fz(1,1)-fz2(ii)
                  scalz22=(xv81*azx(ii)+yv81*azy(ii)+zv81*azz(ii))
                  scalx22=(xv81*axx(ii)+yv81*axy(ii)+zv81*axz(ii))
                  scaly22=(xv81*ayx(ii)+yv81*ayy(ii)+zv81*ayz(ii))
      r2=sqrt((xv78-scalx22)**2+(yv78-scaly22)**2+(zv78-scalz22)**2)

         if(r2.gt.dist)then
      fvalue=fvalue+10*(r2-dist)**2
         endif
         if(r2.lt.distmin)then
      fvalue=fvalue+10*(r2-distmin)**2
         endif
         enddo
999     continue

        if(flag.eq.1)print*,fvalue
      RETURN
      END
```