University of Florence

International Doctorate in Structural Biology

Cycle XXIII (2008-2010)

NMR studies on the structures and dynamics of calcium binding proteins

Ph.D. thesis of

Xiaoyu Hu

Tutor	Coordinator
Prof. Giacomo Parigi	Prof. Ivano Bertini

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

Proteins are very important macromolecules and perform crucial functions in all biological processes. They function as catalysts, transport and store other molecules such as oxygen, provide mechanical support and immune protection, generate movement, transmit nerve impulses, and control growth and differentiation [1]. Large numbers of proteins perform their functions with the cooperation of metal ions, like troponin and calmodulin with calcium [2] - [7], Cytochrome C oxidase and Sco with copper [8] [9], hemoglobin and ferritin with iron [10] [11]. This thesis is focused on calcium binding proteins.

1.1 Calcium in biological systems

Calcium is the most abundant inorganic element in biological systems, and accounts for about 2% (1400 gram) of the adult human body weight. About 99% of calcium in the physique is in skeleton and teeth as hydroxyapatite. The remaining 1% circulates in blood, extracellular spaces or is stored intracellularly in distinct organelle like ER/SR (endoplasmic/sarcoplasmic reticulum). [6] [12]

During evolution, Ca^{2+} has emerged as the most versatile intracellular messenger that mediates a wide range of biological processes, like for example, muscle contraction, secretion, glycolysis and gluconeogenesis, ion transport and cell division [13]. Within a typical cell, the intracellular calcium concentration $[Ca^{2+}]_i$ is around 100 nM, sizably lower than the extracellular level $[Ca^{2+}]_o$ by approximate 12,000-fold. This gradient is maintained through various plasma membrane calcium pumps.

External biochemical or biophysical signals, such as hormones, neurotransmitters, electrical impulses or light, can often lead to transient increases of 10–100 folds in $[Ca^{2+}]_i$ [14]. This increase of $[Ca^{2+}]_i$ is the result of the influx of the extracellular Ca^{2+} through three types of gated channel (by ligand, Ca^{2+} or voltage) located on the membrane and the release of calcium from internal stores to the cytosol. Then calcium ions bind particular proteins, which contribute to switch the cell from an “off” state to an “on” state and perform many cellular processes [15] [16]. However, the amount of calcium ion in cells is highly regulated by several mechanisms in order to protect the cell against toxic effects under long-term exposure to high $[Ca^{2+}]_i$ (>5-10uM): Ca^{2+} ions are actively pumped outside the cell across the membrane or pumped into ER/SR through the plasma membrane calcium ATPase (PMCA pump) and

the sacro(endo)plasmic reticulum calcium ATPase (SERCA pump), separately; Ca^{2+} can also be expelled from the cell by Na^{+}/Ca^{2+} exchanger (NCX), which has lower affinity but higher capacity with respect to ATPases (Figure 1.1)[12]. After that, the $[Ca^{2+}]_i$ come back to the normal level and the cell switches to the "off" state again.

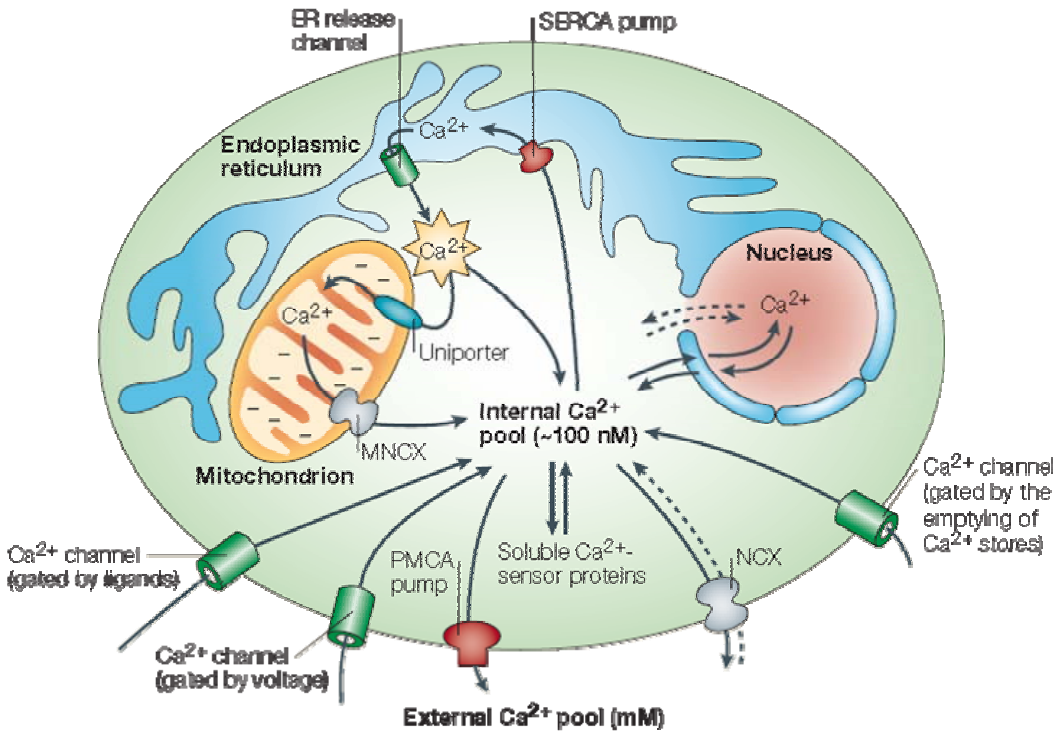


Figure 1.1: The control of cellular Ca^{2+} (reprint from Carafoli, E. (2003) Nature Reviews Molecular Cell Biology, 4:326-332.)

Calcium ions are also known to play very important roles outside cells. Normal $[Ca^{2+}]_o$ is necessary to maintain the rhythmic contractions of heart and the stability of mammalian neural tissues[17][18].

1.2 Calcium binding EF-hand proteins

During the "on" cell stage, calcium-binding proteins participate in calcium cell signalling pathways by binding to Ca^{2+}. In a large number of these calcium-binding proteins, a "helix-loop-helix" secondary structure domain can be found, which is termed as "EF-hand motif". [19]

Functionally, calcium binding EF-hand proteins can be divided into two main subsets: the Ca^{2+} sensors and the Ca^{2+} buffers. The Ca^{2+} sensors transduce Ca^{2+} signals into metabolic or mechanical responses by interacting with various targets. This interaction is accomplished

predominantly through a Ca^{2+}-induced conformational change and exposure of a hydrophobic target interaction site, which is sequestered in the apo form. Calmodulin (CaM), recoverin and some S100 proteins are the classic examples of Ca^{2+} sensors. On the other hand, signal modulators such as calbindin D9K and parvalbumin, known as Ca^{2+} buffer, do not require exposure of hydrophobic target interface to function. This kind of EF-hand proteins functions by modulating the Ca^{2+} signal transduction or removing this potentially harmful ion from the cytoplasm. [20]- [22]

Structurally, EF-hand motifs are divided into two major groups: the canonical EF-hands, like calmodulin (CaM) and the prokaryotic CaM-like protein calerythrin, and the pseudo EF-hands exclusively found in the N-terminal of S100 and S100-like proteins. The canonical EF-hand motif is usually characterized by a sequence of 12 residues with the pattern X*Y*Z*#*–X**–Z, where X, Y, Z, #, –X and –Z are the ligands that participate in metal coordination and the stars represent intervening residues. This sequence forms a loop that can accommodate calcium with distinct geometries: seven ligands coordinate calcium at the vertices of a pentagonal bipyramid. The residue at the -X axis coordinates the Ca^{2+} ion through a bridged water molecule (Figure 1.2). The residues at position 1 and 12 are usually bidentate ligands (Asp or Glu) at axis X and –Z, respectively. These residues are negatively charged and will make a charge-interaction with the positively charged calcium ion. The sixth residue is necessarily Gly due to the requirements of the backbone conformation. The remaining residues are usually hydrophobic which can form a hydrophobic core and stabilize the two helices. The pseudo EF-hand motif is generally constituted by 14 residues which chelate Ca^{2+} primarily via backbone carbonyls in the positions 1, 4, 6, 9 and via the side chain oxygen groups of the residue in the position 14, which is usually Glu (Figure 1.3). [23] [24]

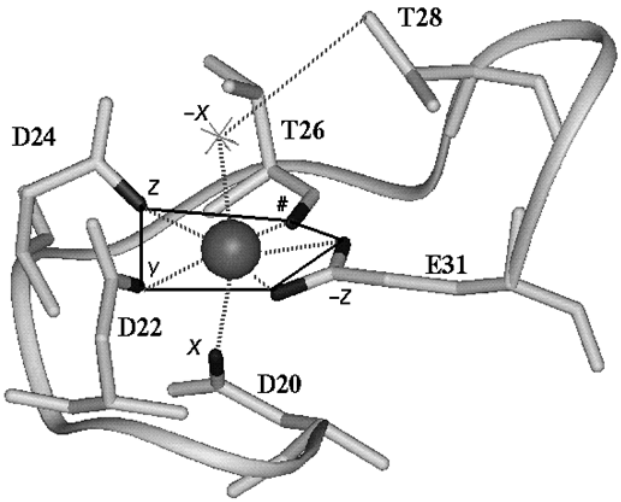


Figure 1.2: Calcium coordination in canonical EF-hand motif (first loop of calmodulin). Oxygen atoms ligating calcium are colored black. The water ligand is shown as a star. The five ligands that form the base of the pentagonal bipyramid are connected with black lines. The extensive hydrogen bonding patterns found in the loop are shown in broken lines (PDB code ICLL). (reprint from Nelson MR, Chazin WJ. (1998) L.J. Van Eldik and D.M. Watterson, eds. Academic Press, San Diego, 17-64.)

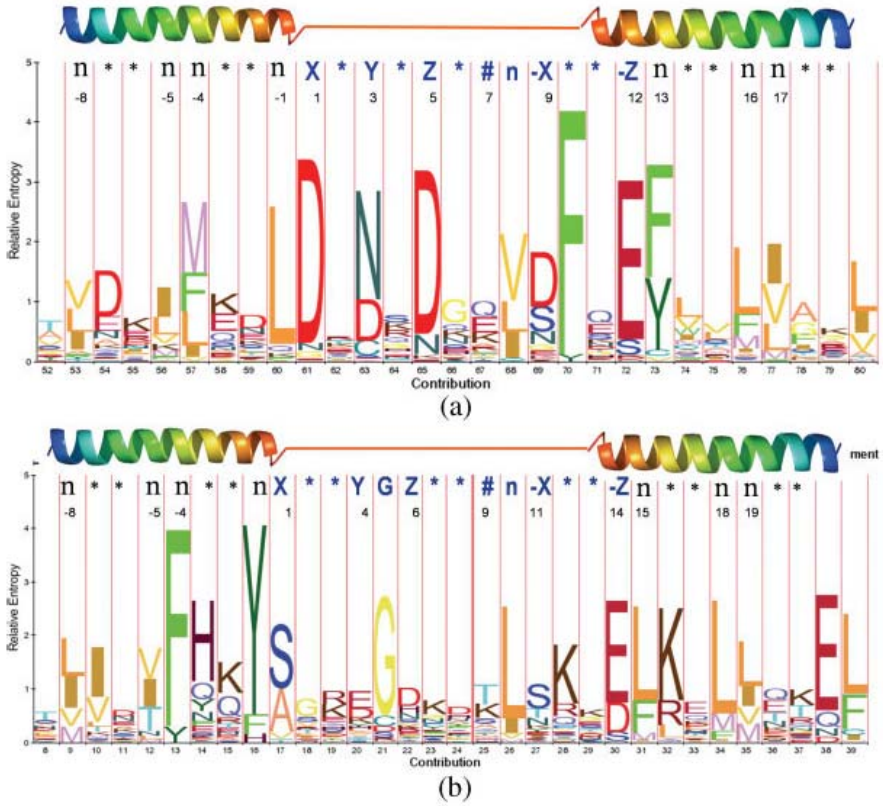


Figure 1.3: Consensus sequence of canonical EF-hand (a) and pseudo Ef-hand domains (b).
n: the hydrophobic residues within the flanking helices.
#: the potential Ca^{2+} binding ligands involving the backbone carbonyl groups.
(reprint from Zhou Y, Wang W, Kirberger M, Lee HW, Ayalasomayajula G, Yang JJ. (2006) Proteins, 65: 643-655.)

1.3 Structural and functional features of S100 protein

The S100 protein family is composed by low molecular weight (10-12 KDa) proteins, and represents the largest subfamily of the EF-hand calcium binding proteins. More than 20 S100 members have been found so far. At least 16 S100 genes are clustered on human chromosome 1q21, with few exceptions of 4p16 (S100P), 5q14 (S100Z), 21q22 (S100B) and xp22 (Calbindin D9K). [25] The expression of S100 genes are ubiquitous, but tissue- and cell type-specific for individual S100 gene. The S100 proteins have 25-65% identity at the amino acid level and contain two EF-hand motifs flanked by conserved hydrophobic residues and separated by a linker region. The C-terminal domain EF-hand motif is a 12-residue canonical EF-hand with higher calcium binding affinity, whereas the N-terminal domain EF-hand motif is a 14-residue pseudo EF-hand motif, also known as S100-specific EF-hand, with a lower calcium binding affinity [26]. The sequences of the linker region and the C-terminal extension are the most variable among the S100 proteins, and are believed to be the active binding site to particular targets. [27]

Except calbindin D9k, which is monomeric, all other S100 proteins exist within cells as dimers (homodimers or heterodimers). Helix I and helix IV and the linker loop of each subunit form an X-type four-helix bundle which is held together by non-covalent bonds. They present the major contributors to the dimeric interface.

S100 proteins undergo large conformational change upon calcium binding. Helix III, which is almost antiparallel to helix IV in apo form, rearranges itself by rotating of around 40-50°, and becomes perpendicular to helix IV after Ca^{2+}-binding. The hinge loop between helix II and III swings out, and a cleft is formed in each monomer. As a result, some hydrophobic residues, which are usually buried in apo S100 protein, are exposed to the protein surface. (Figure 1.4) These conformational changes are important for the Ca^{2+}-dependent interaction with target proteins [28].

S100A7, which does not bind Ca^{2+} in the N-terminal EF-hand due to the lack of the Glu residue in position 14, also has a similar conformational rearrangement upon Ca^{2+}-binding [29]. This indicates that the calcium-dependent conformational change is largely determined by the C-terminal EF-hand calcium binding. The calcium load of N-terminal calcium binding

site, instead, only causes minor alterations of its backbone conformation. S100A10, another special member of the S100 family, has "calcium bound" conformation with hydrophobic residues exposed, even in the absence of calcium ions. S100A10 lost the ability to bind Ca^{2+}, because the first putative binding loop lacks three residues[30], and some amino acid replacements in the second putative binding loop (Asp-Cys at position 61, Glu-Ser at position 70 with respect to calbindin D_{9k}) hamper the ability of this loop to bind Ca^{2+} [31][32]. Furthermore, the affinity for calcium in S100A3 is so low ($K_d \approx 20$ mM) that calcium binding is actually prevented in vivo. [33]

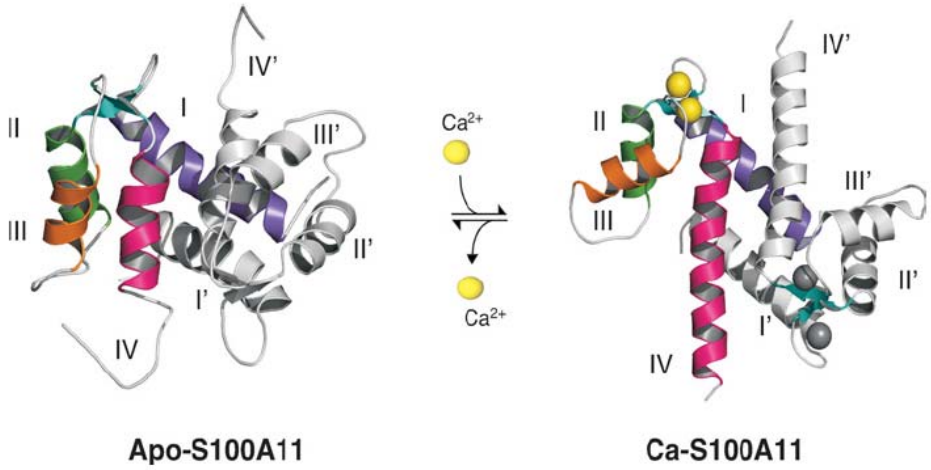


Figure 1.4: Calcium-dependent conformational change in S100 proteins. (reprint from Santamaria-Kisiel L, Rintala-Dempsey AC, Shaw GS. (2006) Biochem J, 396:201-214.)

Besides Ca^{2+} binding, the binding of Zn^{2+} has been proven in some S100 proteins (S100B, S100A2, S100A3, S100A7 and S100A12) with a wide range of affinity (from K_d=4 nM to 2 mM) [34]-[38]. However, S100 binding of Zn^{2+} in the cytoplasm is rather unlikely, because of its sub-nanomolar intracellular concentration. On the other hand, several S100 proteins have been also found in the extracellular space, where Zn^{2+} concentration can be sizably larger [39], and the Zn^{2+} binding could be biologically meaningful. The Zn^{2+} binding sites, distinct from those of Ca^{2+}, are located at the subunit–subunit interface, and comprise His and Glu residues from the helix IV and the N-terminal Ca^{2+} binding loop of the other subunit [40]. Zn^{2+}-binding regulates the calcium binding affinity positively (S100B, S100A12) or negatively (S100A2), and can lead to similar conformational changes as Ca^{2+}-binding with different amplitudes. This suggests that Zn^{2+}-binding may modulate the Ca^{2+}-dependent target interactions or in some cases promote interactions with different targets. [41] [42]

Cu^{2+}-binding was also observed in a few S100 proteins (S100B, S100A5, S100A12 and S100A13) [43]-[46]. S100B dimer binds 4 Cu^{2+} with a K_d of 0.46 uM, most of which can be replaced by Zn^{2+}, suggesting that Cu^{2+} share the same binding site with Zn^{2+} in S100B. However, different S100 proteins could have various Cu^{2+} binding properties. Unlike Ca^{2+} and Zn^{2+} bindings, no protein conformational change was reported upon Cu^{2+} binding of S100 proteins [41] [45].

Within cells, S100 proteins are involved in the regulation of protein phosphorylation, some enzyme activities, cell growth and differentiation, Ca^{2+} homeostasis and dynamics of cytoskeleton components. Both S100A4 and S100B inhibit p53 phosphorylation leading to inhibition of its transcriptional activity, thereby compromising p53 tumor-suppressor activity [47] [48]. S100A10 inhibits ANXA2 (a Ca^{2+}-dependent phospholipids-, membrane-, and cytoskeleton-binding protein) phosphorylation as a result of the sequestration of ANXA2 in the cytoplasm, and consequently S100A10 modulates the activities of ANXA2 [49]. S100B and S100A1 stimulate a membrane-bound guanylate cyclase activity in photoreceptor cells and play a role in dark-adaptation of photoreceptors [50]. The expression of S100A8 and S100A9, the formation of the S100A8-S100A9 heterodimer, and the S100A8-S100A9 dependent regulation of casein kinase I and II activity are related to a definite functional stage of macro phages and the occurrence of an inflammatory response [51]. S100A2 might have a tumor suppressor function through promoting p53 transcriptional activity. S100A1 is localized in membranes of the SR (sarcoplasmic reticulum) in striated muscle cells and in the perinuclear region of several cell types, and stimulates Ca^{2+}-induced Ca^{2+} release in skeletal muscle cells. Calbindin D9K functions as Ca^{2+} buffer and modulates the Ca^{2+} homeostasis [52].

In addition to their intracellular functions, several S100 proteins are secreted upon Ca^{2+} signaling via vesicle fusion with the cell membrane or some other mechanisms into the extracellular space, where they stimulate neuronal survival and/or differentiation and astrocyte proliferation, cause neuronal death via apoptosis, and modulate the activity of inflammatory cells (Figure 1.5). Interaction of S100B with the receptor for advanced glycation end products (RAGE) has been implicated in both neurotrophic and neurotoxic effects of S100B via activation of different pathways.

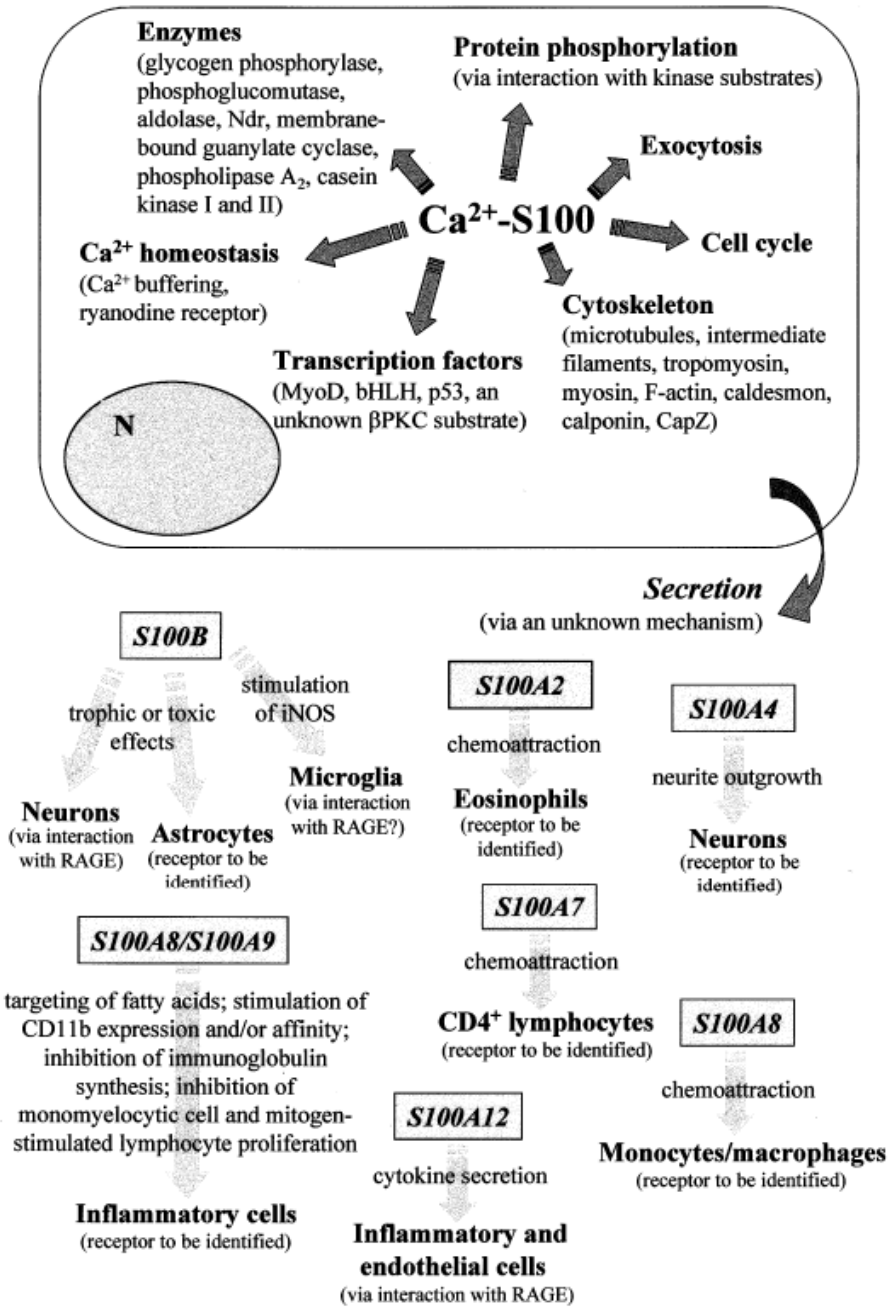


Figure 1.5: Schematic representation of intracellular and extracellular roles of S100 proteins. Not all the targets/activities putatively regulated by S100 proteins are in a Ca2+-dependent manner (N denotes the nucleus). (reprint from [53]Donato R. (2001), Int J Biochem Cell Biol. 33: 637-68.)

S100A12 binds RAGE on endothelial cells, mononuclear phagocytes and lymphocytes, and triggers cellular activations (secretion of IL-1β and TNF-α) in a RAGE-mediated manner. S100A8-S100A9 complex is secreted by neutrophils by a secretion pathway that depends on an intact microtubule (MT) network. The heterocomplex was shown to recruit monocytes to inflammatory sites by enhancing CD11b expression and/or its affinity in human monocytes, and participate in the transendothelial migration mechanism. The extracellular effects of

S100A1, S100A2, S100A4, S100A6, S100A7, S100A10 and S100P have also been described [54]-[59] but only the secretion of the three proteins mentioned above have been documented [52].

The chromosome 1q21, where most S100 genes are located, is structurally conserved during evolution. Within this chromosomal region, several rearrangements which occurred during tumor development have been described [41] [60]. This might be the reason of the deregulatory expression of the S100 gene in various tumor cells. S100 gene deregulatory expressions have also been detected in several pathological conditions: cardiomyopathies, neurodegenerative and inflammatory disorders [60].

S100A5 is a S100 member but poorly characterized at the protein level. Immunohistochemical analysis indicates that it is expressed in very restricted regions of the adult brain (the olfactory bulb, the brainstem, and the spinal trigeminal tract). Flow dialysis revealed that the S100A5 binds four Ca^{2+} ions per dimer with strong positive cooperativity and an affinity 20–100 folds higher than the other S100 proteins studied under identical conditions. It is also reported that S100A5 binds two Zn^{2+} ions and four Cu^{2+} ions per dimer. Although the structures and functional role of some other S100 proteins have been characterized, no structural data was available for S100A5. Also there are only few published works on the dynamic properties of S100 proteins, which could provide important features for ligand binding. Regarding the biological role that S100A5 could play, only one article reported that totally resected WHO grade I meningiomas did not recur or recurred later with high levels of S100A5 than with low S100A5 levels. [61] - [63]

S100A16 is a novel member of the S100 family. It is widely distributed in humans, and highly conserved in mammals. Highest S100A16 mRNA levels were found in the esophagus followed by adipose and colon, low levels were found in lung, brain, pancreas, and skeletal muscle; S100A16 mRNA expression was up-regulated in tumors of bladder, lung, thyroid gland, pancreas, and ovary. Furthermore, investigation of S100A16 intracellular localization in human glioblastoma cells revealed an accumulation of the protein within nucleoli and a translocation to the cytoplasm in response to calcium stimulation. S100A16 presents uncommon characteristics with respect to other S100 proteins. The N-terminal pseudo EF-hand was predicted to be functionally inactive since it comprises 15 amino acids, and

lacks the conserved Glu residue at the last position, analogously to S100A7. However, the role that S100A16 plays inside cells and/or in the extracellular space is not clear. [64] [65]

1.4 CaM and its inter-domain motion

Calmodulin (CaM) is a well known EF-hand Ca^{2+}- binding protein expressed in all eukaryotic cells. It can bind to and regulate a large number of protein targets, thereby affecting many different cellular functions, i.e. metabolism, cell proliferation, cytoskeletal dynamics, cell–cell interaction and development (Figure 1.6) [66] [67].

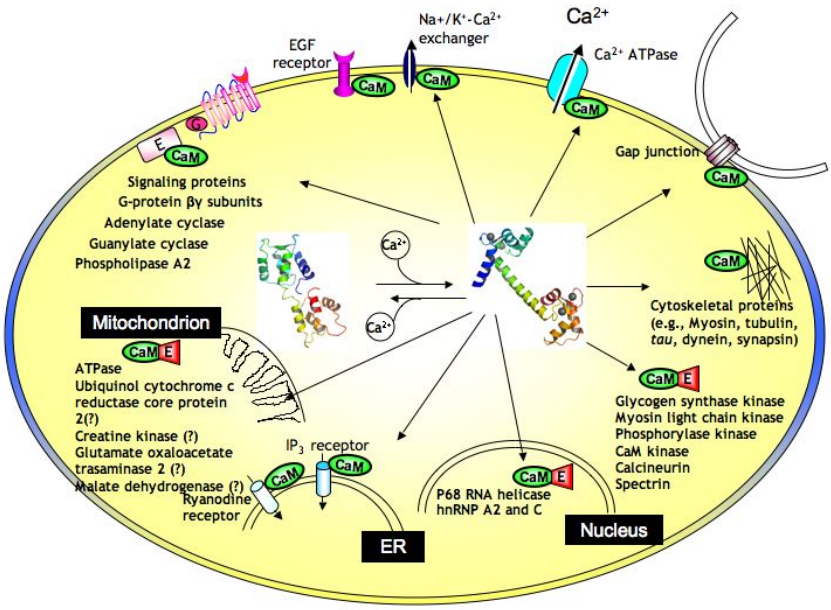


Figure 1.6: Cellular distribution of CaM binding proteins at different cellular compartments. (reprint from the website of the Department of Chemistry, Georgia State University, USA http://chemistry.gsu.edu/faculty/Yang/Signaling.htm)

Many of the proteins that CaM binds are unable to bind calciums, themselves. This makes CaM essential as a calcium sensor and signal transducer. When the cell is excited by external signals, the $[Ca^{2+}]_I$ increases and calcium binds CaM. This causes CaM to be released from neuromodulin or neurogranin, which have much more affinity with apo-CaM than with Ca^{2+}-CaM. In a Ca^{2+}-dependent manner, CaM can interact with at least 30 different enzymes and proteins including Ca^{2+}-transport ATPase, phosphodiesterase, myosin light chain kinase (MLCK) and other CaM-dependent protein kinases, calcineurin, and nitric oxide synthas (NOS). Apo CaM also play important roles to maintain cellular normal activities by binding to various actin-binding proteins (e.g. myosins), cytoskeletal and membrane proteins (e.g. neuromodulin and neurogranin), enzymes (e.g. phosphorylase b

kinase and iNOS) and channels and receptors (e.g. SR Ca^{2+} release channel and inositol 1,4,5-trisphosphate receptor) [68].

The CaM structure has been well characterized both in the apo and calcium form. The protein consists of two similar globular domains, the N-terminal (1-77) and the C-terminal (81-148) domains, each containing two canonical EF-hand motifs which allow CaM to bind up to four calcium ions per molecule. The two domains are connected by a very flexible central linker in solution, so that the relative orientation of the two domains can change readily. Within each of the two domains of CaM, significant conformational changes occur upon calcium binding. In the absence of calcium each EF-hand motif adopts a "closed conformation", with the helices in an almost antiparallel arrangement and most of the hydrophobic residues shielded from the solvent (Figure 1.7 A). Binding of Ca^{2+} normally causes a rearrangement of the helices to the "open conformation", with many hydrophobic residues exposed on the surface of the protein and a large hydrophobic surface created on each domain (Figure 1.7 B). This conformational switch allows CaM to bind to target proteins through these hydrophobic surfaces [69]-[71]. Upon target binding, the inter-domain motion of CaM may be lost. As shown in Figure 1.7 C, the two domains of CaM wrap around the bound peptide (CaM kinase II CaM binding domain), form stable hydrophobic interface and make the CaM conformation rigid.

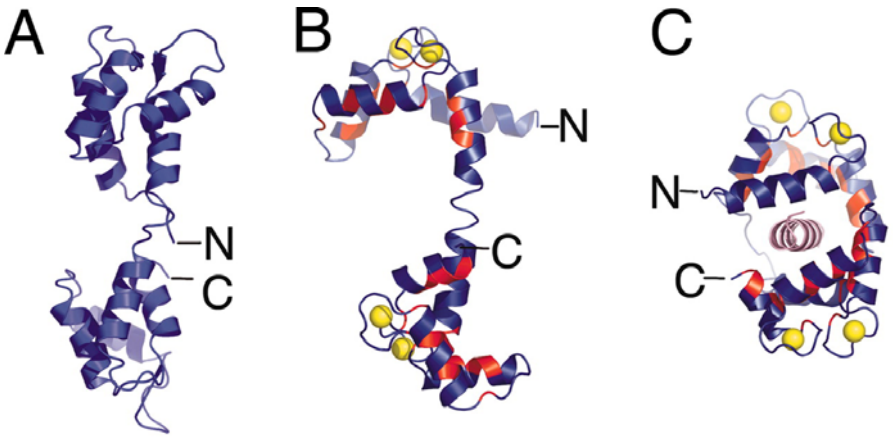


Figure 1.7: CaM conformation. A: apo-CaM (1CFD); B: Ca^{2+} bound CaM (3CLN), the hydrophobic patches important for target binding are shown in red and calcium atoms are shown in yellow spheres; C: complex structure of Ca^{2+}-CaM bound to CaM kinase II peptide (CaM binding domain) (1CM1). (reprint from Shifman JM, Choi MH, Mihalas S, Mayo SL, Kennedy MB. (2006), Proc Natl Acad Sci USA, 103: 13968-73.)

The inter-domain motion of CaM in solution can be maintained in CaM-target

complexes if the target interacts with only one domain of CaM, e.g. in the CaM-C20W complex [72]. Even if CaM binds to targets with both N- and C- domains, in particular cases, inter-domain motions are still observable [73] [74]. It is a significant challenge to understand the relative positions to some extent that the domains can experience, the relative weight of each conformation and the time scale of the motions involved. X-ray crystallography can not give us such information because usually only a single "frozen" conformational state is present at the solid state and the existence of inter molecular packing forces may easily alter the behavior of proteins in solution [75] [76]. NMR relaxation parameters can provide the information on local and global motions of CaM but not the details about the interesting inter-domain motions between the two assumed internally rigid domains.

Paramagnetic restraints obtained from NMR experiments have been proved to be powerful tools to study protein-protein and protein-ligand interactions and for protein structure determination. Self-orientation residual dipolar coupling (rdc), which are distance-independent, provide information of the orientation of molecular nuclear pairs with respect to magnetic susceptibility tensor, and can also be used to study the inter-domain motion within the CaM molecule [77]. The pseudocontact chemical shift (pcs) is a contribution to chemical shift of the nucleus caused by the presence of centers with unpaired electrons. This perturbation is measurable when the magnitude of the magnetic moment of the unpaired electron depends on the molecular orientation with respect to the magnetic field vector. Magnitude of the pcs is proportional to r^{-3} [78], where r is the distance between the free radical center and the nucleus. This inverse third power dependence allows observation of pcs for larger distance ranges up to approximately 40 Å [79]. Pcs can provide long distance restraints in structure determination as NOEs and another important tool for the CaM inter-domain motion study.

From 2003, N60D CaM has been used in CERM to study the inter-domain motions in both free and target-bound states by substituting the Ca^{2+} ion on the second calcium binding site with a paramagnetic lanthanide. Protocols have been developed to calculate the maximum allowed probability (MAP) of each possible conformation. Based on these methods, the conformations with highest MAP values of free CaM, CaM complexes with α-Synuclein (AS), Death-associated protein kinase (DAPk), DAPK-related protein 1 (DRP1)

and myelin basic protein (MBP) have been determined or refined from X-ray structures [80]-[82].

1.5 hMBF1: a reported CaM binding protein

Human MBF1 (hMBF1) is a 16 KDa polypeptide, also known as endothelial differentiation-related factor (EDF) 1. It is a transcriptional cofactor that mediates transactivation by stabilizing the protein-DNA interaction. Structural data available on hMBF1 show that the protein contains an N-terminal flexible region and a C-terminal well structured domain. It was reported that hMBF1 is a CaM target protein, and that their interaction is regulated by the level of Ca^{2+}. [83]- [85] In order to acquire more structural information of the interaction, NMR experiments were performed.

1.6 Aims of the research

During the three years of the PhD course, my research was focused on the determination of the structural and dynamic characterization of Ca^{2+} binding proteins by NMR spectroscopy.

As mentioned before, S100 family is a big protein family. Since the first two S100 proteins (S100A1 and S100B) were isolated from bovine in 1965, more than 20 members have been found up to now. Although most of S100 proteins have similar structural features, their properties of metal binding, dynamics, expression, target recognition and biological function vary significantly. In the last three decades, S100 proteins have received increasing attention due to its various physiological and pathological characters, such as cancer and cardiomyopathies. Even so, their protein targets, intracellular transport and secretion mechanisms, regulation factors and solution dynamics are still largely unclear. In my research, the structural and dynamic features of S100A5 and S100A16 have been studied in both apo and Ca^{2+} bound forms. The conformational changes that have been observed upon calcium binding were compared with those of the other S100 proteins. Furthermore, the binding of Zn^{2+} and Cu^{2+} ions have also been investigated. These information are likely helpful to predict and understand some of their possible functions.

N60D mutant CaM was previously studied in CERM to investigate inter-domain motions. This was achieved by substituting a paramagnetic lanthanide in the second calcium

binding site of N-terminal domain. This single mutant strategy was successful for CaM, but can not be used for other multidomain proteins which do not bind metal ions or have not a mutant able to bind paramagnetic ions selectively in a single site. In order to solve this problem, a small lanthanide probe ClaNP-5 (Caged Lanthanide NMR Probe 5) was used to bind CaM (H107C, N111C) with two disulfide bonds. The probe has already been proved rigid with respect to the protein backbone [86]. Refined MAP program was used to calculate the conformations with largest MAP values in agreement with the paramagnetic-based restraints observed for the lanthanide tagged protein.

Reference List

(1) Jeremy M Berg, John L Tymoczko, and Lubert Stryer Biochemistry, 5th edition, part 1 chapter 3.

(2) Hartshorne DT, Mueller H. (1968) Biochem. Biophys. Res. Commun. 31: 647-653.

(3) Schaub M C, Perry SV. (1969) Biochem. J. 115: 993-1004.

(4) Wilkinson J M, Perry SV, Cole HA, Trayer I P. (1971) Biochlem. J. 124: 55-56.

(5) Wilkinson J M, Perry SV, Cole HA, Trayer I P. (1972) Blochem. J. 127: 215-228.

(6) Carafoli E. (2002) Proc. Natl. Acad. Sci. USA 99: 1115–1122.

(7) Kawasaki K, Kretsinger RH. (1994) Protein 1: 343–517.

(8) Abriata LA, Banci L, Bertini I, Ciofi-Baffoni S, Gkazonis P, Spyroulias GA, Vila AJ, Wang S. (2008) Nat Chem Biol. 4(10):599-601.

(9) Banci L, Bertini I, Cavallaro G, Rosato A. (2007) J Proteome Res. 6(4):1568-79.

(10) Goswami T, Rolfs A, Hediger MA (2002) Biochem Cell Biol. 80(5):679-89.

(11) Liu X., Theil EC. (2005).Acc. Chem. Res. 38: 167–175.

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

(13) Carafoli E, Santella L, Branca D, Brini M. (2001) Crit Rev Biochem Mol Biol. 36(2):107-260.

(14) Balla T. (2009) Cell Calcium.45 (6): 527-534.

(15) Berridge MJ. (2001) Novartis Found Symp. 239:52-64.

(16) Berridge MJ, Bootman MD, Roderick HL. (2003) Nat Rev Mol Cell Biol, 4(7):517-529.

(17) Egelman DM, Montague PR. (1999) Biophysical Journal, 76:1856-1867.

(18) Orlov SN, Aksentsev SL, Kotelevtsev SV. (2005) Cell calcium, 38:53-57.

(19) Nakayama S, Kretsinger RH, (1994) Annual Review of Biophysics and Biomolecular Structure, 23:473-507.

(20) Nelson MR, Chazin WJ. (1998) BioMetals, 11:297-318.

(21) Grabarek Z. (2006) Journal of Molecular Biology, 359:509-525.

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

(23) Lewit-Bentley A, Rety S. (2000) Current Opinion in structure Biology, 10:637-643.

(24) Zhou Y, Wang W, Kirberger M, Lee HW, Ayalasomayajula G, Yang JJ. (2006) Proteins, 65: 643-655.

(25) Schäfer BW, Heizmann CW. (1996) Trends Biochem. Sci, 21:134–140.

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

(27) Santamaria-Kisiel L, Rintala-Dempsey AC, Shaw GS. (2006) Biochem J, 396:201-214.

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

(29) Brodersen DE, Etzerodt M, Madsen P, Celis JE, Thøgersen HC, Nyborg J, Kjeldgaard M (1998) Structure, 6:477-489.

(30) GerK V, Weber K. (1985) EMBO J, 4:2917-2920.

(31) Rety S, Sopkova J, Renouard M, Osterloh D, Gerke V, Tabaries S, Russo-Marie F, Lewit-Bentley A (1999) Nat Struct Biol, 6:89-95.

(32) Kube E, Becker T, Weber K, Gerke V (1992) J Biol Chem, 267:14175-14182.

(33) Fritz G, Heizmann CW, Kroneck PM. (1998) Biochim Biophys Acta, 1448:264–276.

(34) Wilder PT, Baldisseri DM, Udan R, Vallely KM, Weber DJ. (2003) Biochemistry, 42:13410-21.

(35) Randazzo A, Acklin C, Schäfer BW, Heizmann CW, Chazin WJ. (2001) Biochem Biophys Res Commun, 288:462-467.

(36) Föhr U, Heizmann CW, Engelkamp D, Schäfer BW, Cox JA. (1995) J Biol Chem 270:21056–61.

(37) Brodersen DE, Nyborg J, Kjeldgaard M. (1999) Biochemistry, 38:1695-1704.

(38) Moroz OV, Burkitt W, Wittkowski H, He W, Ianoul A, Novitskaya V, Xie J, Polyakova O, Lednev IK, Shekhtman A, Derrick PJ, Bjoerk P, Foell D, Bronstein IB (2009) BMC Biochem 10:11

(39) Hwang JJ, Park MH, Choi SY, Koh JY. (2005) J Biol Chem, 280:11995-12001.

(40) Moroz OV, Wilson KS, Bronstein IB. (2010) Amino Acids, published online.

(41) Heizmann CW, Cox JA. (1998) BioMetals, 11:383-397.

(42) Franz C, Durussel I, CoxJA, Schäfer BW, Heizmann CW. (1998) J Biol Chem, 273:18826-34.

(43) Nishikawa T, Lee IS, Shiraishi N, Ishikawa T, Ohta Y, Nishikimi M. (1997) J Biol Chem. 272:23037-41.

(44) Schäfer BW, Fritschy JM, Murmann P, Troxler H, Durussel I, Heizmann CW, Cox JA. (2000) J Biol Chem. 275:30623-30.

(45) Moroz OV, Antson AA, Grist SJ, Maitland NJ, Dodson GG, Wilson KS, Lukanidin E, Bronstein IB. (2003) Acta Cryst, D59:859-67.

(46) Arnesano F, Banci L, Bertini I, Fantoni A, Tenori L, Viezzoli MS. (2005) Angew Chem Int Ed Engl, 44:6341-4.

(47) Lin J, Blake M, Tang C, Zimmer D, Rustandi RR, Weber DJ, Carrier F. (2001) J Biol Chem 276:35037-41.

(48) Grigorian M, Andresen S, Tulchinsky E, Kriajevska M, Carlberg C, Kruse C, Cohn M, Ambartsumian N, Christensen A, Selivanova G, Lukanidin E. (2001), J Biol Chem 276: 22699–22708.

(49) Gerke V, Moss SE. (2002),Physiol Rev 82:331–371.

(50) Rambotti MG, Giambanco I, Spreca A, Donato R. (1999), Neuroscience 92:1089–1101.

(51) Murao S, Collart FR, Huberman E. (1989). J Biol Chem, 264:8356–8360.

(52) Donato R. (2003), Microsc Res Tech 60:540-551.

(53) Donato R. (2001), Int J Biochem Cell Biol. 33: 637-68.

(54) Huttunen HJ, Kuja-Panula J, Sorci G, Agneletti AL, Donato R, Rauvala H. (2000), J Biol Chem 275:40096–40105.

(55) Nagy N, Brenner C, Markadieu N, Chaboteaux C, Camby I, Schäfer BW, Pochet R, Heizmann CW, Salmon I, Kiss R, Decaestecker C. (2001), Lab Invest 81:599–612.

(56) Novitskaya V, Grigorian M, Kriajevska M, Tarabykina S, Bronstein I, Berezin V, Bock E, Lukanidin E. (2000), J Biol Chem 275:41278–41286.

(57) Donato R. (2007),Curr Mol Med. 7:711-24.

(58) Tan JQ, Vorum H, Larsen CG, Madsen P, Rasmussen HH, Gesser B, Etzerodt M, Honore B, Celis JE, Therstrup-Pederson K. (1996), J Invest Dermatol, 107:5–10.

(59) Kassam G, Le B-H, Choi K-S, Kang H-M, Fitzpatrick SL, Louie P, Waisman DM. (1998), Biochemistry 37:16958–16966.

(60) Heizmann CW, Fritz G, Schäfer BW. (2002), Frontiers in Bioscience, 7: D1356-D1368.

(61) Wicki R, Schäfer BW, Erne P, Heizmann CW. (1996), Biochem Biophys Res Commun, 227: 594-599.

(62) Schäfer BW, Fritschy JM, Murmann P, Troxler H, Durussel I, Heizmann CW, Cox JA. (2000), Journal of Biological Chemistry, 275: 30623-30630.

(63) Hancq S, Salmon I, Brotchi J, De Witte O, Gabius HJ, Heizmann CW, Kiss R, Decaestecker C (2004) Neuropathol Appl Neurobiol, 30:178–187.

(64) Marenholz I, Heizmann CW. (2004) Biochem Biophys Res Commun, 313:237-244.

(65) Sturchler E, Cox JA, Durussel I, Weibel M, Heizmann CW. (2006), J Biol Chem, 281: 38905-17.

(66) Zhang M, Yuan T. (1998), Biochem Cell Biol, 76: 313-323.

(67) Huang Y, Zhou Y, Yang W, Butters R, Lee HW, Li S, Castiblanco A, Brown EM, Yang JJ. (2007), J Biol Chem, 282:19000-10.

(68) Jurado LA, Chockalingam PS, Jarrett HW. (1999), Physiological Reviews, 79: 661-682.

(69) Capozzi F, Luchinat C, Micheletti C, Pontiggia F, Proteome J. (2007), Res, 6:4245-4255.

(70) Yamniuk AP, Vogel HJ. (2004), Mol Biotechnol, 27: 33-57.

(71) Shifman JM, Choi MH, Mihalas S, Mayo SL, Kennedy MB. (2006), Proc Natl Acad Sci USA, 103: 13968-73.

(72) Elshorst B, Hennig M, Forsterling H, Diener A, Maurer M, Schulte P, Schwalbe H, Griesinger C, Krebs JF, Schmid H, Carafoli E. (1999), Biochemistry, 38: 12320-32.

(73) Maximciuc AA, Putkey JA, Shamoo Y, MacKenzie KR. (2006),Structure, 14: 1547-1556.

(74) Bertini I, Gupta YK, Luchinat C, Parigi G, Peana M, Sgheri L, Yuan J. (2007), J Am Chem Soc, 129:12786-94.

(75) Chou JJ, Li S, Klee CB, Bax A. (2001), Nature Struct.Biol, 8: 990-997.

(76) Goto NK, Skrynnikov NR, Dahlquist FW, Kay LE. (2001), J.Mol.Biol, 308: 745-764.

(77) Bertini I, Janik MB, Liu G, Luchinat C, Rosato A. (2001), J Magn Reson. 148:23-30.

(78) McConnell HM, Robertson RE. (1958), J Chemical Physics 29:1361.

(79) Allegrozzi M, Bertini I, Janik MBL, Lee YM, Liu G, Luchinat C. (2000), J Am Chem Soc, 122:4154-61.

(80) Bertini I, Gupta YK, Luchinat C, Parigi G, Peana M, Sgheri L, Yuan J. (2007), J Am Chem Soc, 129: 12786-94.

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

(82) Bertini I, Kursula P, Luchinat C, Parigi G, Vahokoski J, Wilmanns M, Yuan J. (2009), J Am Chem Soc, 131: 5134-44.

(83) De Koning B, Blombach F, Wu H, Brouns SJ, van der Oost J. (2009), Biochem Soc Trans, 37:52-57.

(84) Takemaru K, Harashima S, Ueda H, Hirose S. (1998), Mol Cell Biol, 18:4971-6.

(85) Mariotti M, De Benedictis L, Avon E, Maier JA. (2000), J Biol Chem, 275: 24047-51.

(86) Keizers PH, Saragliadis A, Hiruma Y, Overhand M, Ubbink M. (2008), J Am Chem Soc, 130: 14802-12.

2
METHODOLOGY

Nuclear magnetic resonance (NMR) spectroscopy is a tool of structural biology used to obtain information on the structure and the dynamics of biological molecules in solution. It is applied for the study of small molecular compounds such as ethanol and dextrose as well as macro molecular compounds such as proteins and nucleic acids. The development of pulsed Fourier transform NMR spectroscopy by Ernst and Anderson [(1)] and the design of multidimensional NMR spectroscopy by Jeener[(2)] make this technique increasingly widely applied. Up to date, NMR spectroscopy and X-ray crystallography are the only techniques capable of determining the three dimensional structures of macromolecules at atomic resolution. However, 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. In addition, NMR spectroscopy is a unique technique for investigating time-dependent chemical phenomena, including reaction kinetics and intramolecular dynamics. Recently, paramagnetic restraints pseudocontact shifts (pcs) and self-orientation residual dipolar couplings (rdc) have been used as long-range structural restrains for structure determination and for the study of the inter-domain motions within multi-domain proteins in solution[(3)(4)]. My three years' PhD researches are focused on the structural determination and the characterization of the dynamics of two S100 proteins, S100A5 and S100A16, and on the investigation of the conformations sampled by CaM in solution by using paramagnetic constraints. NMR titration studies were also performed to detect the possible interaction of CaM with target proteins, like hMBF1.

2.1 Structure determination by NMR spectroscopy

Canonical protocols for NMR structure determination include the protein solution preparation, the collection and analysis of NMR datasets and the structural calculations based on NMR data.

Typically, uniformly ^{13}C and ^{15}N labeled proteins are used for structure determination experiments. The double labeled protein makes it possible to record an experiment that transfers magnetization over the peptide bond, and thus connect different spin systems through bonds. The resonance signals of backbone nuclei are assigned by using a number of 3D experiments, e.g. HNCO, HNCA, HNCACB and CBCACONH. All these experiments

consist of a ^{1}H-^{15}NHSQC plan expanded with a carbon dimension: HNCO contains the chemical shift of each HSQC peak and the carbonyl carbon of its previous residue; HNCA provides the chemical shifts of the alpha carbons of each HSQC peak and the sequentially previous one; HNCACB and CBCACONH contain both the alpha and the beta carbon signals. And then backbone sequential assignment can be undertaken by matching the shifts of each spin system and those of the previous one. When the backbone assignment has been obtained, usually, it is possible to assign the signals of side chain carbons and protons by using the HCCH-TOSCY spectrum, which is basically a TOCSY experiment resolved in an additional carbon dimension.

The nuclear Overhause effects (NOE) are due to dipolar interactions between different nuclei. The intensity of NOE, i.e. the volume of the corresponding cross peak in a NOESY spectrum, is related to the product of the inverse sixth power of the internuclear distance *r* and an effective correlation time τ_c. NOEs between pairs of hydrogen atoms are traditionally of prime interest for structural studies [(5)].

$$NOE \propto \frac{1}{\langle r \rangle^{6}} \cdot f(\tau_c) \qquad [1]$$

Since also affected by intramolecular motions, NOE data are usually treated as upper limit (UPL) of the inter-atomic distances instead of precise distance restraints. In practice, a ^{1}H-^{1}H NOE is observed between protons which are separated by less than 5-6 Å, and with a lower limit of 2.0 Å, which represents the sum of the van der Waals radii of the two NOE-connected hydrogen atoms. In addition to distance restraints, torsion angle restraints are also used in structural determination, which will improve the quality of the structure. Torsion angles can not be acquired from NMR spectra directly. However, torsion angles can be predicted by using several programs, such as TALOS and CSI, based on a complete set of backbone chemical shifts of H_N, N, H_α, C_α, C_β, and CO resonances. In my researches, TALOS and its upgraded version TALOS+ have been used for the prediction of torsion angle restraints [(6) (7)]. By using both distance and angle restrains, the solution structures of S100A5 and S100A16 were calculated by using the program CYANA 2.1[(8) (9)] and minimized with the program AMBER 10[(10)]. The qualities of the structures were evaluated with the programs PROCHECK-NMR [(11)] and WHATIF [(12)].

2.2 Dynamic properties characterization by NMR relaxation parameters

Since the 1970s NMR has been used to study the interplay between biomolecular structures, dynamics and functions [13]. It is well known that there is an intimate relation between protein dynamics and molecular function. For example, protein dynamics contribute to the thermodynamic stability of functional states and play an important role in catalysis, ligand binding, molecular recognition processes and allostery. NMR spectroscopy is the unique method suitable to study many of these dynamic processes, because information can be obtained for motions that span a wide range of time scales such as fast internal motions (ps-ns) and slow motions (μs-ms). Passing from the protein overall correlation time in the order of 10-100 μs order. [14]- [16]

Relaxation of an amide ^{15}N nucleus spin at high field is dominated by the dipolar interaction with the directly attached proton spin and by the chemical shift anisotropy (CSA) mechanism [17]. In a ^{15}N relaxation experiment, a non-equilibrium spin order is created, and spin evolution is recorded to follow its relaxation to the equilibrium condition. At equilibrium, the nucleus is in the lower energy level and the ^{15}N magnetization is aligned along the external magnetic field. This alignment can be changed by radio frequency pulses. After absorbing the energy of the pulses, the nucleus will be in the higher energy level and the ^{15}N magnetization is not parallel to the external field any more. The magnetization will relax back to equilibrium, along the direction of the external magnetic field, with a time constant called longitudinal relaxation time T_1. Due to the presence of the angle between ^{15}N magnetization and the external field in non-equilibrium condition, the magnetization vector has a component perpendicular to the external magnetic field. The time constant for this spin component to return to equilibrium is called transverse relaxation time T_2. Another relaxation parameter is the hetero-nuclear NOE, which is measured by saturating the proton (^1H) signal and observing changes in the ^{15}N signal intensities.

The relaxation parameters are related to the spectral density function $J(\omega)$ of the ^1H-^{15}N bond vector by the following equations [18]:

$$T_1^{-1} = R_1 = (d^2/4)[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 J(\omega_N) \qquad [2]$$

$$T_2^{-1} = R_2 = (d^2/8)[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N) + 6J(\omega_H)] + (c^2/6)[4J(0) + J(\omega_N)] + R_{ex} \quad [3]$$

$$NOE = 1 + (d^2/4R_1)(\gamma_H/\gamma_N)[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \quad [4]$$

where

$$d = (\mu_0 h \gamma_H \gamma_N / 8\pi^2)\langle r_{NH}^{-3} \rangle \quad [5]$$

$$c = \omega_N \Delta\sigma / \sqrt{3} \quad [6]$$

In equation [3], it is sometimes necessary to account for chemical exchange effect by adding an exchange contribution $R_{ex}=\omega_N^2\Phi_{ex}$ to the predicted R_2, where Φ_{ex} is a constant that depends on the chemical shift differences, populations, and interconversion rates for the exchanging species (19). In equation [2] to [6], $\mu_0=4\pi \times 10^{-7}$ kg m s^{-2} A^{-2} is the permeability of free space, h=6.6262 × 10^{-34} erg · s is Planck's constant, γ_H and γ_N are the gyromagnetic ratios of ^{1}H and ^{15}N(2.6753 × 10^8 rad · s^{-1} · T^{-1} and -2.71 × 10^7 rad · s^{-1} · T^{-1}, respectively), r_{NH}=1.02 Å is the nitrogen- hydrogen bond length, ω_H and ω_N are Larmor frequencies of ^{1}H and ^{15}N, and Δσ=-160 ppm is the chemical shift anisotropy measured for ^{15}N nuclei in helical polypeptide chains[20]. For a protein in solution, the spectral density function J(ω) depends on the global correlation time (τ_c) and on the internal motion of the ^{1}H-^{15}N bond vector (τ_i). J(ω) can be described by Lipari-Szabo Model-Free formulism [21]:

$$J(\omega) = \frac{2}{5}[S^2 \frac{\tau_c}{1+\omega^2\tau_c^{\,2}} + (1-S^2)\frac{\tau_e}{1+\omega^2\tau_e^{\,2}}] \quad [7]$$

where $\tau_e^{-1} = \tau_c^{-1} + \tau_i^{-1}$. [8]

τ_e is the effective internal correlation time. In the limit $\tau_e << \tau_c$ and $S^2 \approx 1$, Equation [7] can be written as:

$$J(\omega) = \frac{2}{5}[\frac{\tau_c}{1+\omega^2\tau_c^{\,2}}]. \quad [9]$$

In this kind of rigid systems, where the H-N vectors are fixed, the τ_c value of any H-N vector is equal to the molecular overall tumbling time, assuming an isotropic model. The time scale for protein molecular tumbling is usually in the order of 10ns or longer, depending on the size of the protein. If protein backbone amides have conformational dynamics on a time

scale faster than the molecular tumbling, i.e. on the ps-ns time scale, the ^{15}N R_1 values of the corresponding residues are characterized by higher values than the average value, and NOE and R_2 values are lower than the average value. 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. Detailed examples on the characterization of local motions in S100A5 and S100A16 are shown in the Result section 3.1 and 3.2.

2.3 Paramagnetic restraints (pcs and rdc) and their application

As mentioned before, paramagnetic restraints pcs and rdc are increasingly widely used in NMR macromolecular structural and dynamics studies. Paramagnetic lanthanide ions are generally applied as the source of the paramagnetic restraints. The paramagnetic lanthanide ions have variable strength and degree of anisotropy, and this permits to achieve conformational information from the different ions [22] [23].

Pseudocontact shifts (pcs) are related to the magnetic anisotropies and the structural parameters through the following equation:

$$\delta_i^{\mathrm{pcs}} = \frac{1}{12\pi r_i^3}\left[\Delta\chi_{\mathrm{ax}}\left(3\cos^2\vartheta_i - 1\right) + \frac{3}{2}\Delta\chi_{\mathrm{rh}}\sin^2\vartheta_i\cos 2\varphi_i\right] \qquad [10]$$

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:

$$\Delta\chi_{\mathrm{ax}} = \chi_{zz} - \frac{\chi_{xx} + \chi_{yy}}{2} \qquad [11]$$

$$\Delta\chi_{\mathrm{rh}} = \chi_{xx} - \chi_{yy}. \qquad [12]$$

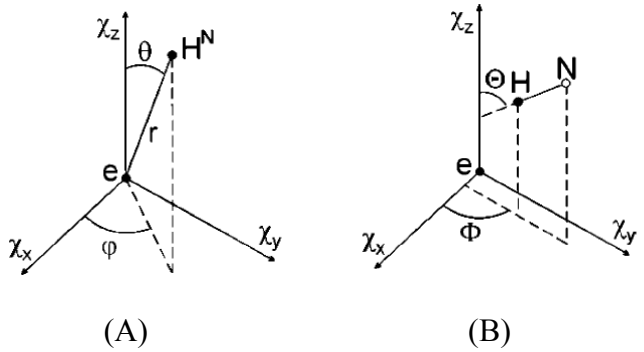
r_i is the distance between the atom i and the metal ion, ϑ_i and φ_i are the spherical angles of atom i with respect to the principal axes of the magnetic susceptibility tensor centered on the metal ion (Figure 2.1 A). Different metal ions induce different pcs values, depending on their magnetic anisotropy. After the complete assignment of the ^{1}H-^{15}N-HSQC spectra of the paramagnetic and the analogous diamagnetic sample, the pcs values of N and ^{N}H of each amino acid are easily obtained from the difference between the corresponding chemical shifts of the paramagnetic and diamagnetic forms. By investigating the pcs of different

lanthanides, the distance and angular information of the nuclei located up to approximately 40 Å from the lanthanide ion can be obtained.[24]- [28].

The residual dipolar coupling (rdc) between two spins occurs if the molecules exhibit a partial alignment in solution, which leads to an incomplete averaging of spatially anisotropic dipolar couplings. When molecules bind paramagnetic metal ions like paramagnetic lanthanides, self-orientation rdc arise as a result of the molecular partial alignment induced by the magnetic susceptibility anisotropy at high magnetic field. Differences in ^{1}J ^{15}N-^{1}H splittings between the paramagnetic and the diamagnetic forms of the protein backbone amides, most frequently used in protein investigations, are described by the equation [28][29]:

$$\Delta\nu_{\mathrm{RDC}}\ (\mathrm{Hz}) = -\frac{1}{4\pi}\frac{B_0^{\ 2}}{15\mathrm{k}T}\frac{\gamma_{\mathrm{N}}\gamma_{\mathrm{H}}\hbar}{2\pi r_{\mathrm{NH}}^{3}}\left[\Delta\chi_{\mathrm{ax}}(3\cos^2\Theta - 1) + \frac{3}{2}\Delta\chi_{\mathrm{rh}}\sin^2\Theta\cos 2\Phi\right] \qquad [13]$$

where Θ is the angle between the N-H vector and the *z* axis of the χ tensor, Φ is the angle between the projection of the N-H vector on the *xy* plane of the χ tensor and the *x* axis (Figure 2.1 B), and $\Delta\chi_{\mathrm{ax}}$ and $\Delta\chi_{\mathrm{rh}}$ are defined the same as in Equation [10]. Unlike NOEs and pcs, rdc values provide orientational information, both short range and long range, instead of distance information between two atoms. Rdc are becoming increasingly important as a powerful complement to NOE in structural determination and a useful tool for protein dynamic investigation [30]. In practical applications, rdc data are usually obtained from IPAP-HSQC experiments as the difference in the doublet splitting in the indirect ^{15}N dimension between the paramagnetic form and diamagnetic form [31].



(A) (B)

Figure 2.1: Schematic representation of the geometric dependence of $^{1}H_N$ pcs (A) and ^{1}H-^{15}N rdc (B). (reprint from Pintacuda G, Keniry MA, Huber T, Park AY, Dixon NE, Otting G. (2004), J Am Chem Doc, 126: 2963-70.)

In this thesis, two kinds of mutated CaM (N60D and H107C, N111C) were used for the study of inter-domain motion: N60D CaM was loaded with Tm^{3+}, Tb^{3+} and Dy^{3+} on the second Ca^{2+} binding site; H107C, N111C CaM was loaded with Lu^{3+}, $Yb^{3+}$and Tm^{3+} through the small tag ClaNP-5 mentioned before. The magnetic susceptibility tensors of the paramagnetic metals were obtained from the pcs values detected on the domain where the Ln^{3+} ion were located using the program FANTASIAN [32]. The relative position between the two domains of CaM is calculated with the pcs and rdc values on the other domain using the program MAP approach [33].

Reference List

(1) Ernst RR, Anderson WA. (1966), Rev Sci Instrum, 37: 93-102.

(2) Jeener J. (1994), Les Editions de Physique, Les Ulis, France, pp1-379.

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

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

(5) Wüthrich K. (2002), Nobel Lecture 235-267.

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

(7) Shen Y, Delaglio F, Cornilescu G, Bax A. (2009), J Biomol NMR, 44:213-223.

(8) Herrmann T, Güntert P, Wüthrich K. (2002), J. Mol. Biol. 319: 209–227.

(9) Guntert P. (2004), Methods Mol Biol, 278: 353-378.

(10) Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister CE, Caldwell JW, Ross WS, and Kollman PA. (2008), Computer program, San Francisco, CA, University of California.

(11) Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM. (1996), J Biomol NMR, 8:477-486.

(12) Vriend G. (1990), J Mol Graphics, 8:52-56.

(13) Wüthrich K, Wagner G. (1978), Trends Biochem Sci, 3:277.

(14) Kay LE, Torchia DA, Bax A. (1989), Biochemistry, 28: 8972.

(15) Palmer AG, Kroenke CD, Loria JP. (2001), Methods Enzymol, 339: 204.

(16) Mittermaier A, Kay LE. (2006), Science, 312: 224-228.

(17) Mandel AM, Akke M, Palmer AG 3rd. (1995), J mol Biol, 246: 144-163.

(18) Abragam A. (1961), Principles of Nuclear Magnetism, pp: 1-599, Clarendon Press, Oxford.

(19) Fischer MWF, Majumdar A, Zuiderweg ERP. (1998), Prog NMR Spectrosc, 33: 207-272.

(20) Hiyama Y, Niu CH, Silerton JV Bavoso A, Torchia DA. (1988), J Am Chem Soc, 110: 2378-83.

(21) Lipari G, Szabo A. (1982), J Am Chem Soc, 104: 4546-58.

(22) Horrocks WD Jr, Sipe JP 3rd. (1972), Science, 177: 994-996.

(23) Hinckley CC. (1969), J Am Chem Soc, 91: 5160-62.

(24) Bertini I, Kursula P, Luchinat C, Parigi G, Vahokoski J, Wilmanns M, Yuan J. (2009), J Am Chem Soc, 131:5134-44.

(25) Bertini I, Luchinat C, Parigi G. (2002), Concepts in Magnetic Resonance, 14:259-286.

(26) Allegrozzi M, Bertini I, Janik MBL, Lee YM, Liu G, Luchinat C. (2000), J Am Chem Soc, 122:4154-61.

(27) Bertini I, Luchinat C, Parigi G, Pierattelli R. (2005), ChemBioChem 6:1536-1549.

(28) Bertini I, Luchinat C, Parigi G. (2002), Progr NMR Spectrosc, 40: 249-273.

(29) Pintacuda G, Keniry MA, Huber T, Park AY, Dixon NE, Otting G. (2004), J Am Chem Doc, 126: 2963-70.

(30) Lipsitz RS, Tjandra N. (2004), Annu Rev Biophys Biomol Struct, 33: 387-413.

(31) Ottiger M, Delaglio F, Bax A. (1998), J Magn Reson, 131: 373-378.

(32) Banci L, Bertini I, Bren KL, Cremonini MA, Gray HB, Luchinat C, Turano P. (1996), J Biol Inorg Chem, 1:117-126.

(33) Bertini I, Gupta YK, Luchinat C, Parigi G, Peana M, Sgheri L, Yuan J. (2007), J Am Chem Soc, 129:12786-94.

3
RESULT

The main achievements that I have obtained during these three years of PhD course are the structure determinations and dynamic characterization of two S100 proteins, S100A5 and S100A16. I have also investigated the presence of interaction between CaM and hMBF1 and the maximum occurrence of different conformations of CaM when the protein is free in solution. All these studies were performed by NMR spectroscopy.

For the structural characterization of S100A5 and S100A16, I performed the NMR experiments, the resonance signals assignments and the structural calculations. I also performed the 15N relaxation measurements in order to obtain information on the protein mobility. The conformational changes upon calcium binding of both the two S100 members were discussed. All these studies are shown in detail in the result section 3.1 and 3.2.

hMBF1 was reported as a CaM target, which interaction is regulated by the level of calcium ions. I studied the *in vitro* interaction between hMBF1 and CaM through NMR titration. These experiments indicated that no interaction is occurring between the two proteins both in absence and in the presence of calcium(II). This is reported in section 3.3.

For studying the conformations possibly experienced by CaM, the maximum occurrence of given conformations can be calculated from the paramagnetic pcs and rdc values. This approach was previously applied to a variant of CaM (N60D), in which paramagnetic lanthanide ions were selectively bound to the second calcium binding site of N-terminal domain. In my study, the same approach has been used with the lanthanide ion bound on the surface of the C-terminal domain of CaM through the tag ClaNP-5. In this way I have obtained new sets of pcs and edc values, averaged according to the sampled protein conformations. By using this new pcs and rdc data together with those arising with the metal in the N-terminal domain, more accurate information can be obtained on the maximum occurrence of the different protein conformations. In this project, I mainly contribute to the NMR spectra acquirement, backbone resonance signals assignments and the calculation of pcs and rdc data. The detailed information is shown in section 3.4.

3.1

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.

J Biol Inorg Chem (2009) 14:1097–1107
DOI 10.1007/s00775-009-0553-1

ORIGINAL PAPER

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

Ivano Bertini · Soumyasri Das Gupta · Xiaoyu Hu · Tilemachos Karavelas · Claudio Luchinat · Giacomo Parigi · Jing Yuan

Received: 20 March 2009 / Accepted: 29 May 2009 / Published online: 18 June 2009

Abstract S100A5 is a calcium binding protein of the S100 family, with one canonical and one S100-specific EF-hand motif per subunit. Although its function is still unknown, it has recently been reported to be one of the S100 proteins able to interact with the receptor for advanced glycation end products. The homodimeric solution structures of S100A5 in both the apo and the calcium(II)-loaded forms have been obtained, and show a conformational rearrangement upon calcium binding. This rearrangement involves, in particular, the hinge loop connecting the N-terminal and the C-terminal EF-hand domains, the reorientation of helix III with respect to helix IV, as common to several S100 proteins, and the elongation of helix IV. The details of the structural changes are important because they must be related to the different functions, still largely unknown, of the different members of the S100 family. For the first time for a full-length S100 protein, relaxation measurements were performed on both the apo and the calcium-bound forms. A quite large mobility was observed in the hinge loop, which is not quenched in the calcium form. The structural differences resulting upon calcium binding change the global shape and the distribution of hydrophobic and charged residues of the S100A5 homodimer in a modest but significantly different manner with respect to the closest homologues S100A4 and S100A6.

Keywords Calcium binding proteins · Calcium-induced conformational rearrangements · EF-hand proteins · Protein dynamics · S100A5

Electronic supplementary material The online version of this article (doi:10.1007/s00775-009-0553-1) contains supplementary material, which is available to authorized users.

I. Bertini (✉) · S. Das Gupta · X. Hu · T. Karavelas · C. Luchinat · G. Parigi · J. Yuan
Magnetic Resonance Center (CERM),
University of Florence,
Via Luigi Sacconi 6,
50019 Sesto Fiorentino, Italy
e-mail: ivanobertini@cerm.unifi.it; bertini@cerm.unifi.it

I. Bertini
Department of Chemistry,
University of Florence,
Via della Lastruccia 3,
50019 Sesto Fiorentino, Italy

Present Address:
T. Karavelas
Department of Chemistry,
University of Ioannina,
45110 Ioannina, Greece

C. Luchinat · G. Parigi
Department of Agricultural Biotechnology,
University of Florence,
via Maragliano 75-77,
50144 Florence, Italy

Introduction

S100 proteins have been found to be implicated in a Ca^{2+}-dependent (and, in some cases, Zn^{2+}- or Cu^{2+}-dependent) regulation of a variety of intracellular and extracellular activities, and several biological targets have been identified for the different proteins [1]. A large variability in the sequence is observed within the protein family, which is responsible for the modulation of the shape and the nature of the binding surface. This modulation is needed to bind different targets, although the overall fold of most members of the family is very similar. The structures available for

S100–target peptide adducts actually display a remarkable lack of uniformity in the orientation of the target [2]. Furthermore, individual S100 proteins can bind different targets in different ways [3].

Several S100 protein structures are available, solved either in solution or in the solid state. Together with S100A4 and S100A6, S100A5, the function of which is still unknown [4], belongs to a well-defined subset of the S100 family showing high homology [5]. The structures of S100A4 and S100A6 are available [6, 7], whereas that of S100A5 is not. We report here the solution structures of S100A5 in both the apo and the calcium(II)-bound states, and compare these structures with those of its closest homologues and of other S100 proteins.

The expression level of S100A5 was demonstrated by immunohistochemical analyses to be restricted to a few specific cells [8], i.e., in the olfactory bulb, in the brainstem, and in the spinal trigeminal tract, analogously to S100A3 and differently from other S100 proteins such as S100A6 and S100B. Marked modifications of the levels of expression of different S100 proteins (including S100A5 and S100A6) occur in connection with the progression of astrocytic tumor malignancy [9]. It was found that totally resected WHO grade I meningiomas with high levels of S100A5 either did not recur or recurred later than those with low S100A5 levels [9]. In this respect, the recently found interaction with the receptor for advanced glycation end products, well known to be involved in tumor outgrowth, may acquire further biological relevance [10].

Most S100 proteins are encoded by genes located in the same chromosome 1q21, with the exception of genes encoding S100B (located on chromosome 21q22), calbindin D_{9k} (also called S100G, located on chromosome Xp22) and S100P (located on chromosome 4p16) [11]. Interestingly, S100A1, S100A3, S1004, S100A5, S100A6, S100A8, S100A9, S100A12, and S100A13 genes are all mapped within a short distance. It is widely appreciated that colocalization of genes may imply coexpression of the proteins [12, 13], and in the case of S100 proteins this probably correlates with the observation of functional heterodimers and possibly with concerted functions [13].

All S100 proteins are constituted by two EF-hand motifs, highly conserved helix–loop–helix structural domains that can each bind a calcium(II) ion. 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 [1, 4, 14]. The N-terminal EF-hand comprises helix I, the S100-specific calcium binding loop I, and helix II, which is separated by a flexible linker, called "hinge loop," from the C-terminal EF-hand, which comprises helix III, calcium binding loop II, and helix IV.

Calcium(II) binding is an important mechanism in cells because 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 [15]. Furthermore, calcium(II) ions play a central role in cell signaling. Calcium(II) binding to EF-hand proteins in fact induces in most cases conformational changes that correlate with binding of target proteins/enzymes involved in a wide variety of cellular processes. The helices in the EF-hand motifs can have an almost antiparallel arrangement, called "closed conformation," or an almost orthogonal arrangement, called "open conformation," depending on the presence of bound calcium [16–22]. The latter conformation exposes large hydrophobic clefts on the protein surface, which acts as a binding region for a variety of targets.

Proteins undergoing changes in the orientation of the helices of each EF-hand motif upon calcium(II) binding are generally functionally related to activation of target proteins, whereas proteins not undergoing conformational changes have the function of calcium buffer and transport [23]. In most cases, S100 proteins undergo smaller structural changes upon calcium(II) binding in the N-terminal domain and larger changes in the C-terminal domain [2], although not as large as those observed for the EF-hand protein calmodulin [24].

All S100 structures determined to date (with the exception of that of calbindin D_{9k}) show that these proteins exist as homodimers, heterodimers, or tetramers [14]. Most of the S100 proteins are homodimers. The dimer interface consists of helices I (I′) and IV (IV′) of each subunit arranged in a X-type four-helix bundle, in both the apo and the calcium-bound states [14]. Calcium binding results in minor alterations of the backbone conformation of calcium binding loop I but causes helix III to reorient and form a more open structure with respect to the apo state. As a result, the hydrophobic residues of helices III and IV in the C-terminal EF-hand are more exposed, thus facilitating the interaction with target proteins. The solution structures obtained here for the apo and calcium(II)-bound forms of S100A5 show that the same behavior applies to this protein, and provide the details of the exposed surface and charge distribution responsible for its possible interactions. Despite the high homology, S100A4, S100A5, and S100A6 show modest but significant differences in the pattern of hydrophobic/hydrophilic/charged residues exposed upon calcium binding. The present data thus provide a further example of the diversity of the exposed protein surface, which is likely to be reflected in a diversity in target proteins.

Materials and 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 *Escherichia coli* BL21 Gold cells. For the production of ^{15}N-labeled or ^{13}C- and ^{15}N-labeled S100A5, cultures were grown in minimal medium using ^{13}C-glucose and/or ^{15}N-ammonium sulfate as the sole carbon and nitrogen source, respectively. Cells were grown at 310 K to an optical density at 600 nm of 0.6 and growth was induced with 1 mM isopropyl β-D-thiogalactopyranoside. After induction, the temperature was reduced to 298 K and the culture was grown overnight. Cells were harvested by centrifugation at 15,000*g* for 15 min and resuspended in lysis buffer [20 mM tris(hydroxymethyl)aminomethane (Tris)–HCl pH 7.5, 1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, 2 mM dithiothreitol (DTT)]. Cell lysis was performed by sonicating with eight bursts of 30 s each. The suspension was ultracentrifuged at 200,000*g* for 30 min. The cleared lysate was precipitated by slowly adding ammonium sulfate to 30% and centrifuging at 15,000*g* for 20 min. 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 eluate was concentrated and applied to a Superdex 75 column equilibrated with 30 mM 2-morpholinoethanesulfonic acid, pH 6.5, 100 mM NaCl, 5 mM DTT. The fractions containing S100A5 were pooled and washed with excess EDTA to remove all metal ions. The yield of S100A5 was 20 mg L^{-1} of culture.

NMR spectroscopy and structure determination

All NMR experiments for assignments were performed at 298 K using a Bruker 500 MHz spectrometer equipped with a cryoprobe. Apo and Ca_2-S100A5 samples (0.4 mM) were ^{13}C- and ^{15}N-labeled, in 30 mM 2-morpholinoethanesulfonic acid, 100 mM NaCl, and 5 mM DTT 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 total correlation spectroscopy, HBHA(CBCACO)HN together with ^{13}C nuclear Overhauser effect spectroscopy (NOESY) heteronuclear single quantum coherence (HSQC), and ^{15}N-NOESY-HSQC experiments. Proton–proton distance restraints were derived from the analysis of 2D NOESY, ^{15}N-NOESY-HSQC, and ^{13}C-NOESY-HSQC spectra acquired using a Bruker 900 MHz spectrometer equipped with a cryoprobe. The spectra were processed using TOPSPIN and analyzed with CARA [25]. The secondary structure elements were predicted from the chemical shift index and the backbone dihedral angles were obtained from TALOS [26], accordingly. The structures were calculated using the program CYANA-2.0 [27]. The two subunits in the dimeric structure were linked together through a chain of dummy atoms with zero van der Waals radii. 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 radii, 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 included for this purpose. 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 structures of the CYANA family were then subjected to restrained energy minimization with AMBER 10 [28]. Nuclear Overhauser effect (NOE) and torsion angle restraints were applied with force constants of 50 kcal mol^{-1} $Å^{-2}$ and 32 kcal mol^{-1} rad^{-2}, respectively. The program PROCHECK-NMR [29] 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 with a 700 MHz spectrometer using standard pulse sequences [30, 31], 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 program [32]. 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. 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 reorientation time were then calculated with the model-free approach [33]. Theoretical predictions of ^{N}H, R_1, and R_2 values for apo-S100A5 and Ca_2-S100A5 were calculated by using the HYDRONMR software program [34].

Metal binding detection

The binding of apo-S100A5 to Ca^{2+} was monitored by following the changes in the chemical shifts of the protein

NMR peaks in the ^{1}H–^{15}N-HSQC spectra upon titration of the apoprotein with calcium ions. The chemical shift perturbation between the free and bound states was obtained for each residue by calculating the composite chemical shifts according to Eq. 1:

$$\delta(\mathrm{HN}) = \sqrt{\frac{\Delta\delta_{\mathrm{H}}^2 + (\Delta\delta_{\mathrm{N}}/5)^2}{2}}, \quad (1)$$

where $\Delta\delta_{\mathrm{H}}$ and $\Delta\delta_{\mathrm{N}}$ are the differences in chemical shifts between the bound and free states of the amide protons and of the nitrogen atoms, respectively.

Accession numbers

Atomic coordinates, structural restraints, and resonance assignments of apo-S100A5 and Ca_2-S100A5 have been deposited in the Protein Data Bank (codes 2KAX and 2KAY) and BioMagResBank (codes 16033 and 16034).

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 with the exception of Cys-43 in the calcium form. The corresponding assignments are deposited in BioMagResBank together with the ^{1}H–^{13}C–^{15}N assignments of the side chain resonances. The types of NMR spectra used for the assignments are described in "Materials and methods".

Ca^{2+} titration of apo-S100A5

The binding of calcium(II) to apo-S100A5 was monitored by following the changes in the ^{1}H–^{15}N-HSQC NMR spectrum of ^{15}N-labeled apo-S100A5 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 on increasing the Ca^{2+} to apo-S100A5 ratio. When a 2:1 ratio (with respect to the protein subunit concentration) was reached, the original peaks, corresponding to the apo form, disappeared. Such behavior 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 the two forms. Figure 1 shows the chemical shift changes

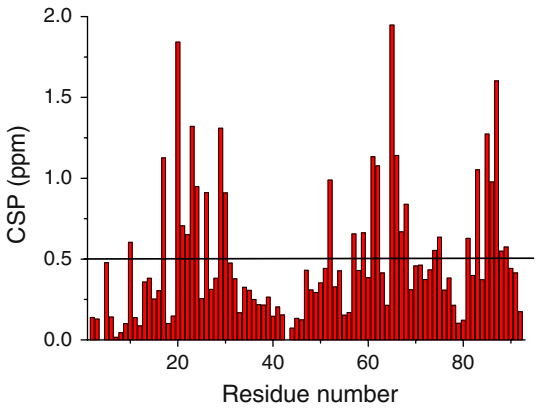


Fig. 1 Composite chemical shift perturbation (*CSP*), as defined in Eq. 1, of apo-S100A5 upon calcium(II) binding. The *horizontal line* indicates the average value

on passing from the apo to the calcium forms of S100A5. The residues undergoing the largest changes in chemical shifts 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 protein, thus indicating that a significant conformational change occurs on passing from the apo form to the calcium form.

The findings of the Ca^{2+} titration experiments are consistent with previous measurements, which provided dissociation constants for the binding of the first and second calcium(II) ions in the submillimolar and submicromolar range, respectively, and a strong positive cooperativity [8]. As already pointed out [8], the affinity of calcium(II) for S100A5 is among the highest in the whole S100 family.

Relaxation measurements

The relaxation parameters for apo and calcium-loaded S100A5 are shown in Fig. 2. Such measurements indicate that the protein is dimeric in both forms. The reorientation 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 agreement with the molecular weight and the reorientation times observed for other S100 homodimers [35–39].

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

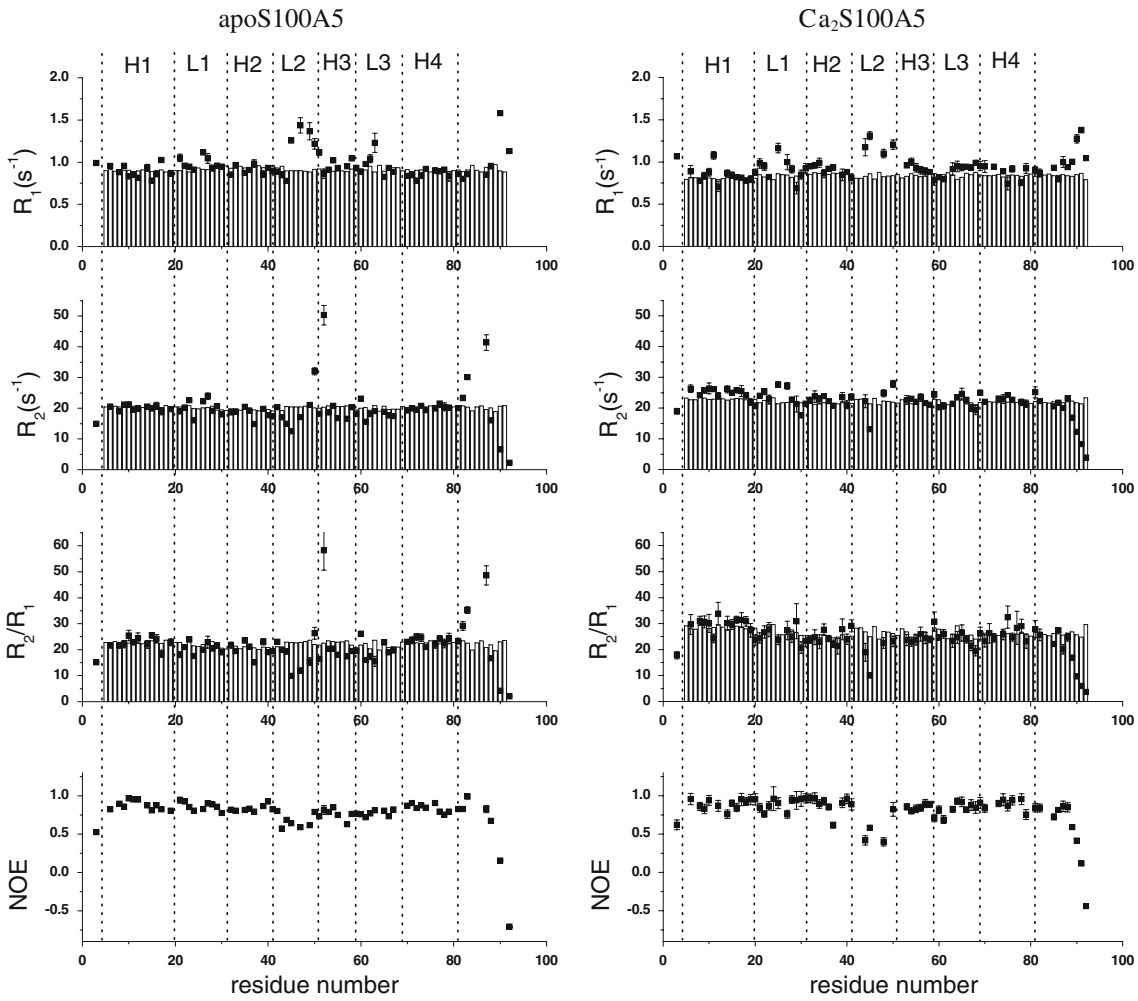


Fig. 2 Sequential plot of the experimental relaxation parameters of apo and Ca^{2+}-S100A5. The values calculated by HYDRONMR are shown as *bars*. *NOE* nuclear Overhauser effect

In the apoprotein, motion on a slower time scale (microseconds to millisecons) was observed for some residues localized at the beginning of helix III (Asp-50 and Ile-52), at the end of helix IV (Tyr-83), and at the C-terminus (Phe-87), as indicated by the significantly larger R_2 value (for Ile-52, in particular, R_2 is 52 s^{-1} compared with an average value of about 20 s^{-1} observed for helix residues). This mobility may originate from backbone amide conformational exchange and/or side-chain rotation. Since the time scale for the conformational exchange (microseconds to milliseconds) is sizably shorter than the mixing time (100 ms) of the NOESY spectra, the conformational reorientations of these residues may cause ^{1}H–^{1}H-NOEs to be observed between nuclei of the side chains of such residues and residues located in quite different positions (see later). Upon calcium binding, residues in the slow motion regime are not observed any more. However, calcium binding does not reduce the fast motion detected for residues in the hinge loop, the observed ^{1}H–^{15}N-NOE being even smaller than in the apo form (the average ^{1}H–^{15}N-NOE in the hinge loop is 0.60 and 0.40 for the apo and calcium forms, respectively).

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 2,752 and 2,530 meaningful upper distance limits per dimer, including 184 and 190 intersubunit upper distance limits for the apo and the calcium forms, respectively, were used (Tables 1, 2). Few NOE patterns were detected for residues in the hinge loop and at the C-terminus, consistent with the observed mobility of such regions. In the calcium form, the Ca^{2+} ions were restrained to be within 3 Å from the oxygen ligand atoms (O of Ser-20, Glu-23, Ser-25, Thr-28 and OE1, OE2 of Glu-33 for the first binding site, and OD1 of

Table 1 Structural restraints and statistical analysis of apo-S100A5

Structural restraints		
NOE upper distance limits		
Intrasubunit	1,284	
Intraresidue	358	
Interresidue		
Sequential ($\|i - j\| = 1$)	395	
Medium range ($\|i - j\| < 4$)	320	
Long range ($\|i - j\| > 5$)	211	
Intersubunit	184	
Dihedral angle restraints		
φ	96	
ψ	96	
Statistical analysis		
	Family	Mean
RMS violations per meaningful distance restraints (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 restraints (°)		
φ	4.87 ± 1.23	4.00
ψ	4.04 ± 1.42	2.82
Average number of restraints per residue	14.96	14.96
Average number of violations per conformer		
φ	6.93 ± 2.00	8.00
ψ	6.63 ± 2.20	4.00
NOE violations between 0.1 and 0.3 Å	9.40 ± 2.59	8.0
NOE violations larger than 0.3 Å	0	0
Average RMSD from 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		
Residues in most favorable regions (%)	82.7[a]/88.1[b]	92.6
Residues in allowed regions (%)	13.2[a]/10.0[b]	6.6
Residues in generously allowed regions (%)	2.2[a]/0.9[b]	0.0
Residues in disallowed regions (%)	1.9[a]/1.0[b]	0.8

NOE nuclear Overhauser effect, *RMS* root mean square, *RMSD* root mean square deviation

[a] RMS deviation values were calculated in the sequence range 3–82

[b] RMS deviation values were calculated excluding flexible loop 41–52 of both subunits

Table 2 Structural restraints and statistical analysis of Ca_2-S100A5

Structural restraints		
NOE upper distance limits		
Intrasubunit	1,170	
Intraresidue	380	
Interresidue		
Sequential ($\|i - j\| = 1$)	325	
Medium range ($\|i - j\| < 4$)	280	
Long range ($\|i - j\| > 5$)	185	
Intersubunit	190	
Dihedral angle restraints		
φ	120	
ψ	120	
Statistical analysis		
	Family	Mean
RMS violations per meaningful distance restraints (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 restraints (°)		
Phi	1.97 ± 0.23	1.6246
Psi	0.63 ± 0.42	0.5218
Average number of restraints per residue	13.75	13.75
Average number of violations per conformer		
φ	9.74 ± 1.76	9.0
ψ	1.83 ± 1.23	2.0
NOE violations between 0.1 and 0.3 Å	4.03 ± 1.93	6.0
NOE violations larger than 0.3 Å	0	0
Average RMSD from 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		
Residues in most favorable regions (%)	86.0[a]/90.4[b]	90.7
Residues in allowed regions (%)	11.2[a]/7.5[b]	9.3
Residues in generously allowed regions (%)	1.9[a]/1.5[b]	0.0
Residues in disallowed regions (%)	0.9[a]/0.6[b]	0.0

[a] RMSD values were calculated in the sequence range 3–90

[b] RMSD values were calculated excluding flexible loop 41–49 of both subunits

Asp-60, Asn-62, OD1, OD2 of Asp-64, O of Glu-66, and OE1, OE2 of Glu-71 for the second binding site). No symmetry constraint was used.

The calculated families of structures are shown in Fig. 3. In both forms, the eight individual helices of the two EF-hand motifs of each subunit present in the dimeric structure are very well defined; the four calcium binding loops are less well defined, whereas the linker regions

between the two EF-hand calcium binding domains are poorly defined. These results are in line with what was previously found for other EF-hand proteins [36, 40]. Each calcium binding loop contains a short antiparallel β strand.

The root mean square deviation (RMSD) from the mean subunit structure for the structured regions of the protein is 0.73 ± 0.10 Å (backbone) and 1.22 ± 0.09 Å (heavy atoms) for apo-S100A5 (residues 3–40, 53–82) and 0.71 ± 0.09 Å (backbone) and 1.20 ± 0.07 Å (heavy atoms) for Ca_2-S100A5 (residues 3–40, 50–90). The RMSD from the mean dimeric structure for the structured regions of the protein is 0.78 ± 0.09 Å (backbone) and 1.25 ± 0.09 Å (heavy atoms) for apo-S100A5 (residues 3–40, 53–82 of both subunits) and 0.83 ± 0.10 Å (backbone) and 1.29 ± 0.09 Å (heavy atoms) for Ca_2-S100A5 (residues 3–40, 50–90 of both subunits). More than 95% of the 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 analyses of both structures are reported in Tables 1 and 2.

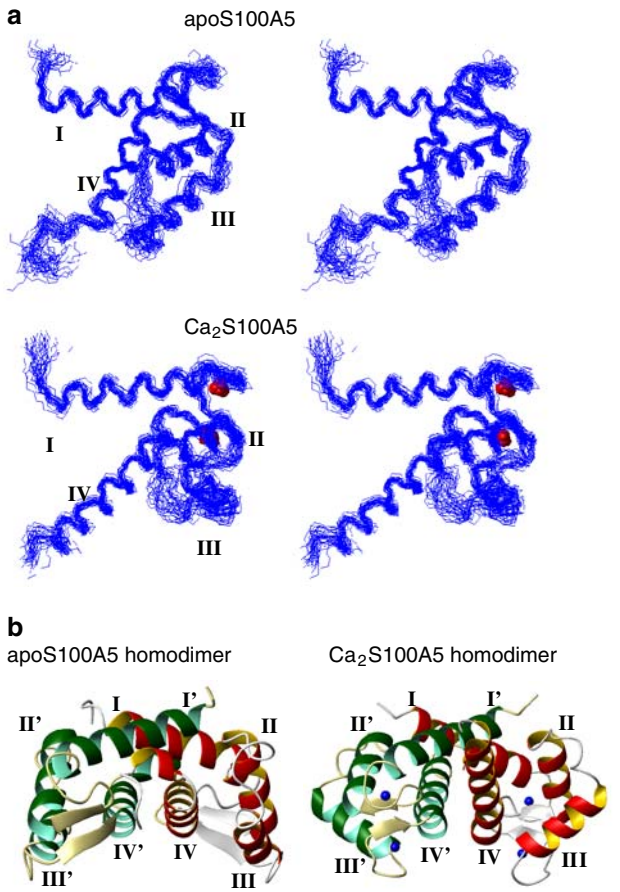


Fig. 3 Stereoview of the families of the solution structures of the S100A5 subunit in the apo and calcium(II)-loaded forms (**a**) and ribbon representation of the homodimer mean structures (**b**) obtained after restrained energy minimization

The relaxation rates were then calculated using HYDRONMR [34] and the minimized mean structures obtained, and are reported in Fig. 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, the differences between calculated and observed values make it easier to appreciate the presence of mobility in some residues of the loops.

In both the apo and the calcium-loaded forms, S100A5 forms homodimers owing to the interactions between helices I and I′ and between helices IV and IV′ of the two subunits. There is a symmetry relationship between the subunits consisting in a twofold rotational axis passing through the dimer interface approximately perpendicular to 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 subunit. Residues Phe-69, Lys-70, Ser-73, and Cys-80 in helix IV also make several contacts with helix I′ and helix IV′ of the other subunit. All these interactions align helix I and helix IV in opposite directions to helix I′ and helix IV′, respectively, in the dimer.

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, such as S100A1, S100A4, S100A6, S100A8, S100A12, S100A13, or S100B [6, 39, 41]. The comparison of the apo and calcium-loaded S100A5 structures shows that the N-terminal EF-hands (residues 5–41) are similar to one another (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 nonnegligible changes in helix IV and in the hinge loop (Fig. 4). These rearrangements upon calcium binding are similar to those observed for other S100 proteins [5, 6, 39, 42, 43], with the exception of S100A10, which is known to have a "calcium-ready state" in both the N-terminal and the C-terminal EF-hands although it does not bind calcium(II) [4]. In apo-S100A6 (1K9P) and apo-S100A13 (1YUR), for instance, helix III is almost antiparallel to helix IV, but opens by 30–40° upon calcium binding (1K9K and 1YUT). The same degree of opening is observed in other EF-hand proteins, such as calmodulin [21], not belonging to the S100 family. In S100A5 the angle between helices III and

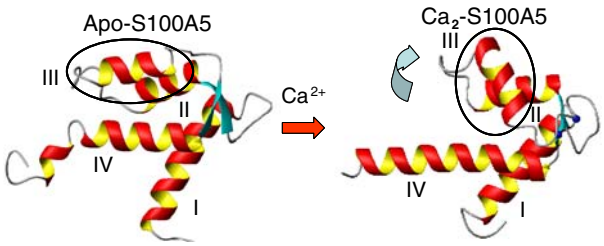


Fig. 4 Major structural differences upon calcium binding: the different angle between helices III (enclosed in *circles*) and IV, and the longer α-helical structure of helix IV

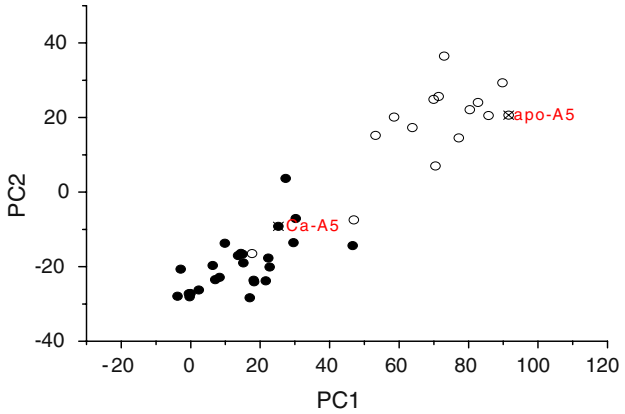


Fig. 5 Principal component plot for the S100 proteins derived from principal components analysis of the six interhelix angles [21]. Apoproteins (S100A1, S100A2, S100A3, S100A4, S100A5, S100A6, S100A10, S100A11, S100A13, S100B, calbindin D_{9k}) are indicated with *open circles* and calcium-loaded proteins (S100A1, S100A4, S100A5, S100A6, S100A7, S100A8, S100A9, S100A12, S100A13, S100B, calbindin D_{9k}, S100P) are indicated with *solid circles*. The two *open symbols* not regularly placed with respect to the others correspond to calbindin D_{9k} and S100A10 in the apo form. The data are based on the structural information reported in the supporting information in [21] as well as on more recent structural information reported in the electronic supplementary material

IV changes from 168° to 118° on passing from the apo to the calcium-bound form, so the two helices are almost perpendicular in the latter form.

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 [21]. 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 us to identify whether S100 proteins have a structure typical of the apo or the calcium-loaded form. The principal component values were thus calculated for the two forms of S100A5, and were plotted together with the values previously calculated for the S100 proteins (Fig. 5), and with the values relative to other S100 proteins deposited in the Protein Data Bank in the meantime, by using the same coefficients for the interhelix angles (calculated with the program MOLMOL) reported by Babini et al. [21]. The figure shows that S100A5 is regularly positioned with respect to the other S100 proteins in both the apo and the calcium-loaded forms, thus pointing to the occurrence of similar structures, and thus of similar overall rearrangement upon calcium(II) binding. It is to be noted that the only two S100 proteins not regularly placed are calbindin D_{9k} and S100A10 in the apo form.

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 result in an increased solvent-exposed surface of the hinge loop and of some positively charged residues of helix II and helix III 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 apo-S100A5, which is loosened upon calcium(II) binding. On the other hand, calcium binding results in a decrease in the exposure of the metal ligand residues Asp-60, Asn-62, Asp-64, and Glu-71 in the C-terminal calcium binding loop. Similarly to what was found for other S100 proteins, 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, where target proteins can be accommodated [14].

In the apo form, residue Ile-52 was identified by relaxation measurements to experience mobility on the millisecond to microsecond time scale. The side chain of this residue experiences NOE contacts with both Met-56 and Tyr-83, which are positioned in opposite directions. As anticipated in "Relaxation measurements," this may be due to the occurrence of conformational exchange, and these data thus indicate that the side chain of Ile-52 can rotate along an axis perpendicular to helix III, so a conformational exchange is also affecting the backbone amide group. As a consequence, the residues forming a hydrophobic patch with Ile-52, and particularly the ones localized on the protein surface such as residue Tyr-83, may also experience sizable mobility on the side-chain and/or backbone atoms.

The slow time scale motion detected for residues Ile-52 and Tyr-83 in the apo form is absent in the calcium form, as a consequence of the conformational rearrangement of both helix III and helix IV. In the apo form, in fact, the

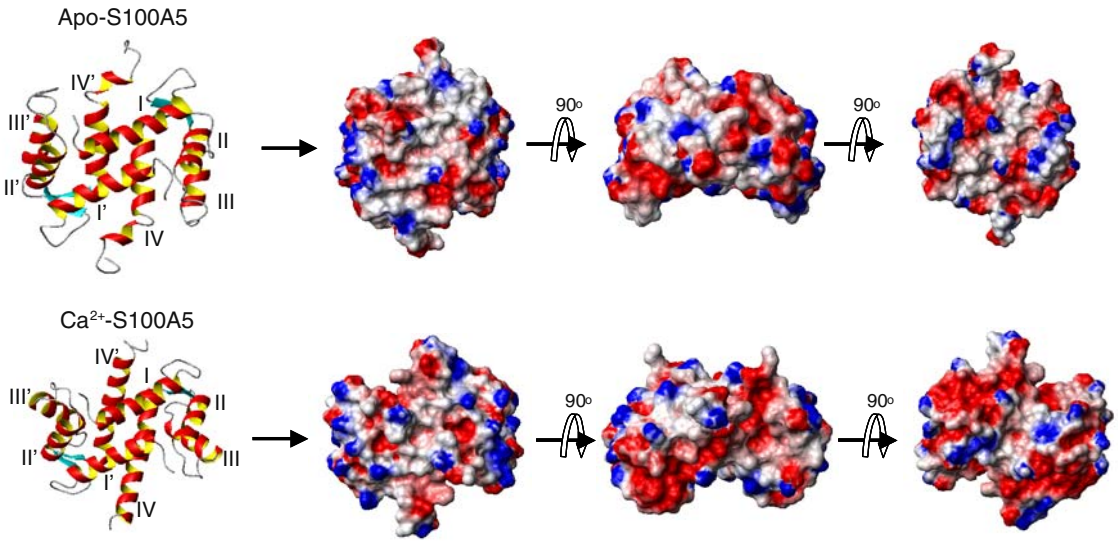


Fig. 6 Electrostatic surface representation of the S100A5 dimer

aromatic ring of Tyr-83 of each subunit participates in a hydrophobic cluster including the side chains of residues Leu-44, Met-47, Ile-52, and Leu-79 of the same subunit. As already noted [6], calcium binding overcomes the hydrophobic interactions that keep this cluster together, so the side chain of Tyr-83 changes orientation and forms new hydrophobic contacts with the side chains of Leu-9, Val-13, and Thr-14 of helix I of the other subunit of the dimeric structure. This rotation is experimentally confirmed by the ^{1}H–^{1}H-NOEs observed between Tyr-83 and residues Leu-44, Lys-48, Ile-52, and Leu-79 of the same subunit in the apo form and with residues Leu-9, Thr-10, and Val-13 of the other subunit 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 deduced for the hinge loop residues from the lower ^{1}H–^{15}N-NOE values measured for the calcium form with respect to 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 [6, 40, 44]. This is due to the unwinding of helix IV in apo-S100A5 at Tyr-83. The different orientation of the side chain of Tyr-83 is in fact responsible for a break in the α-helical structure, being consistent with a regularly formed α helix only in the calcium-loaded form [6].

The hydrophobic residues at the extreme C-terminus (Phe-87 and Leu-88) are important for stabilizing both the apo-S100A5 and the Ca_2-S100A5 homodimer. However, 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 subunit, whereas in the calcium form they have hydrophobic interactions with Val-13 and Thr-14 of helix I of the other subunit.

Analogously to relaxation studies of apo-S100B and apo-S100A4 [35, 37], relaxation studies indicate that helices I and IV are quite rigid, whereas helix III is somewhat more flexible. On the other hand, in S100A5 the loop experiencing a very large 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 [35, 37]. A larger mobility for the hinge loop with respect to the calcium binding loops was also observed for S100A1 [36].

The combination of the structural differences results in a change of the global shape and distribution of surface charges of the S100A5 homodimer upon calcium binding (Fig. 6), whereas no major differences in motion are seen in the two forms. The change in the shape of the protein on passing from the apo to the calcium-loaded form is common to most S100 proteins [14, 39]. On the other hand, the change in the charge distribution seems to depend largely on the particular S100 protein [45]. S100A5 shows a number of charged residues, both positive and negative, on the protein surface in both the apo and the calcium-loaded forms. Interestingly, upon calcium binding, some more exposed positive residues (Lys-48, Lys-57) are moved away from the inner part of the opened cleft, which becomes slightly more hydrophobic, and the negative electrostatic surface is smaller and more clustered around the calcium binding sites. Other S100 proteins show a different change in the surface charge and hydrophobic distribution upon calcium binding: for instance, in S100A6 a larger increase of the hydrophobic surface was observed; in S100B a larger negative charged surface is exposed; in S100A4 the large hydrophobic surface present in the apo form remains exposed also upon calcium binding; in

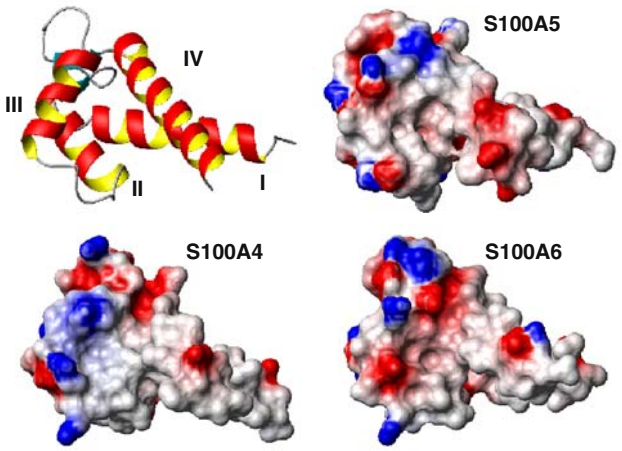


Fig. 7 Electrostatic surface representation of the reported ligand binding area typical of S100 proteins, for S100A4 (2Q91), S100A5 (present work), and S100A6 (1K9K)

S100A13 the charge distribution remains as large as in the apo form, but appears quite different. Overall, the present data provide a further example of the diversity of the exposed protein surface upon calcium(II) binding, which is likely to be reflected in a diversity in target protein(s).

S100A4, S100A5, and S100A6 are the most closely related S100 proteins according to phylogenetic trees constructed on the basis of multiple sequence alignments of S100 proteins [5]. Upon calcium binding, in all these proteins the opened cleft exposes hydrophobic residues, and positive residues become more exposed in the hinge loop and at the end of the third helix, i.e., in the typical protein target binding region. S100A6 also shows some more exposed negative charges at the end of the third helix. The charged groups surrounding the hydrophobic patch, as well as the shape of the surface, are however distinctly different for these proteins in the calcium-bound form (Fig. 7). Such differences observed for S100A4 and S100A6 reflect their different target specificity [46]. This suggests that the function of S100A5 may also be different from that of the other two proteins, and especially from that of S100A6, owing to the different charge pattern.

Acknowledgments This work was supported by Ente Cassa di Risparmio di Firenze, MIUR-FIRB contracts RBLA032ZM7 and RBIP06LSS2, and the European Commission, contracts EU-NMR 026145, SPINE2-COMPLEXES 031220, and LSHG-CT-2004-512052.

References

1. Donato R (1999) Biochim Biophys Acta 1450:191–231
2. Bhattacharya S, Bunick CG, Chazin WJ (2004) Biochim Biophys Acta 1742:69–79
3. Bhattacharya S, Large E, Heizmann CW, Hemmings B, Chazin WJ (2003) Biochemistry 42:14416–14426
4. Santamaria-Kisiel L, Rintala-Dempsey AC, Shaw GS (2006) Biochem J 396:201–214
5. Marenholz I, Heizmann CW, Fritz G (2004) Biochem Biophys Res Commun 322:1111–1122
6. Otterbein L, Kordowska J, Witte-Hoffmann C, Wang CL, Dominguez R (2002) Structure 10:557–567
7. Pathuri P, Vogeley L, Luecke H (2008) J Mol Biol 383:62–77
8. Schäfer BW, Fritschy J-M, Murmann P, Troxler H, Durussel I, Heizmann CW, Cox JA (2000) J Biol Chem 275:30623–30630
9. Hancq S, Salmon I, Brotchi J, De Witte O, Gabius H-J, Heizmann CW, Kiss R, Decaestecker C (2004) Neuropathol Appl Neurobiol 30:178–187
10. Leclerc E, Fritz G, Vetter SW, Heizmann CW (2009) Biochim Biophys Acta 1793:993–1007
11. Ridinger K, Ilg EC, Niggli FK, Heizmann CW, Schäfer BW (1998) Biochim Biophys Acta 1448:254–263
12. Kluger Y, Yu H, Qian J, Gerstein M (2003) BMC Genomics 4:49
13. Mijalski T, Harder A, Halder T, Kersten M, Horsch M, Strom TM, Liebscher HV, Lottspeich F, Hrabe de Angelis M, Beckers J (2005) Proc Natl Acad Sci USA 102:8621–8626
14. Donato R (2001) Int J Biochem Cell Biol 33:637–668
15. Orrenius S, Zhivotovsky B, Nicotera P (2003) Nat Rev Mol Cell Biol 4:552–565
16. Nelson MR, Chazin WJ (1998) Biometals 11:297–318
17. Finn BE, Evenäs J, Drakenberg T, Waltho J, Thulin E, Forsén S (1995) Nat Struct Biol 2:777–783
18. Zhang M, Tanaka T, Ikura M (1995) Nat Struct Biol 2:758–767
19. Kuboniwa H, Tjandra N, Grzesiek S, Ren H, Klee CB, Bax A (1995) Nat Struct Biol 2:768–776
20. Ikura M (1996) Trends Biochem Sci 21:14–17
21. Babini E, Bertini I, Capozzi F, Luchinat C, Quattrone A, Turano M (2005) J Proteome Res 4:1961–1971
22. Capozzi F, Luchinat C, Micheletti C, Pontiggia F (2007) J Proteome Res 6:4245–4255
23. Fragai M, Luchinat C, Parigi G (2006) Acc Chem Res 39:909–917
24. Bertini I, Gupta YK, Luchinat C, Parigi G, Peana M, Sgheri L, Yuan J (2007) J Am Chem Soc 129:12786–12794
25. Keller R (2004) The computer aided resonance assignment tutorial. CANTINA, Goldau
26. Cornilescu G, Delaglio F, Bax A (1999) J Biomol NMR 13:289–302
27. Guntert P (2004) Methods Mol Biol 278:353–378
28. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister CE, Caldwell JW, Ross WS, Kollman PA (2008) AMBER 10. University of California, San Francisco
29. Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM (1996) J Biomol NMR 8:477–486
30. Kay LE, Torchia DA, Bax A (1989) Biochemistry 28:8972–8979
31. Barbato G, Ikura M, Kay LE, Pastor RW, Bax A (1992) Biochemistry 31:5269–5278
32. Goddard TD, Kneller DG (2000) SPARKY 3. University of California, San Francisco
33. Lipari G, Szabo A (1982) J Am Chem Soc 104:4546–4559
34. Garcia de la Torre JG, Huertas ML, Carrasco B (2000) J Magn Reson 147:138–146
35. Inman KG, Baldisseri DM, Miller KE, Weber DJ (2001) Biochemistry 40:3439–3448
36. Zhukov I, Ejchart A, Bierzynski A (2008) Biochemistry 47:640–650
37. Dutta K, Cox CJ, Basavappa R, Pascal SM (2008) Biochemistry 47:7637–7647

38. Bertini I, Fragai M, Luchinat C, Parigi G (2000) Magn Reson Chem 38:543–550
39. Arnesano F, Banci L, Bertini I, Fantoni A, Tenori L, Viezzoli MS (2005) Angew Chem Int Ed 44:6341–6344
40. Smith SP, Shaw GS (1998) Structure 6:211–222
41. Drohat AC, Baldisseri DM, Rustandi RR, Weber DJ (1998) Biochemistry 37:2729–2740
42. Maler L, Sastry M, Chazin WJ (2002) J Mol Biol 317:279–290
43. Bhattacharya S, Chazin WJ (2003) Structure 11:738–739
44. Kilby PM, Van Eldik LJ, Roberts GC (1996) Structure 4:1041–1052
45. Koch M, Diez J, Fritz G (2008) J Mol Biol 378:933–942
46. Gingras AR, Basran J, Prescott A, Kriajevska M, Bagshaw CR, Barsukov IL (2008) FEBS Lett 582:1651–1656

3.2
Structural characterization of human S100A16, a low affinity calcium binder

Elena Babini,[1] Ivano Bertini,[2,3]* Valentina Borsi,[2] Vito Calderone,[2] Xiaoyu Hu,[2] Claudio Luchinat,[2,3] Giacomo Parigi[2,3]

[1]Department of Food Science, University of Bologna, Piazza Goidanich 60, 47521, Cesena, Italy.

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

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

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

J Biol Inorg Chem
DOI 10.1007/s00775-010-0721-3

ORIGINAL PAPER

Structural characterization of human S100A16, a low-affinity calcium binder

Elena Babini · Ivano Bertini · Valentina Borsi · Vito Calderone · Xiaoyu Hu · Claudio Luchinat · Giacomo Parigi

Received: 16 July 2010 / Accepted: 9 October 2010

Abstract The homodimeric structure of human S100A16 in the apo state has been obtained both in the solid state and in solution, resulting in good agreement between the structures with the exception of two loop regions. The homodimeric solution structure of human S100A16 was also calculated in the calcium(II)-bound form. Differently from most S100 proteins, the conformational rearrangement upon calcium binding is minor. This characteristic is likely to be related to the weak binding affinity of the protein for the calcium(II) ions. In turn, this is ascribed to the lack of the glutamate residue at the end of the S100-specific N-domain binding site, which in most S100 proteins provides two important side chain oxygen atoms as calcium(II) ligands. Furthermore, the presence of hydrophobic interactions stronger than for other S100 proteins, present in the closed form of S100A16 between the third and fourth helices, likely make the closed structure of the second EF-hand particularly stable, so even upon calcium(II) binding such a conformation is not disrupted.

Keywords S100A16 · EF-hand proteins · Calcium-binding proteins · S100 proteins · Protein dynamics

An interactive 3D complement page in Proteopedia is available at: http://proteopedia.org/wiki/index.php/Journal:JBIC:3.

Electronic supplementary material The online version of this article (doi:10.1007/s00775-010-0721-3) contains supplementary material, which is available to authorized users.

E. Babini
Department of Food Science,
University of Bologna,
Piazza Goidanich 60,
47521 Cesena, Italy

I. Bertini (✉) · V. Borsi · V. Calderone · X. Hu · C. Luchinat · G. Parigi
Magnetic Resonance Center (CERM),
University of Florence,
Via Luigi Sacconi 6,
50019 Sesto Fiorentino, Italy
e-mail: ivanobertini@cerm.unifi.it

I. Bertini · C. Luchinat · G. Parigi
Department of Chemistry,
University of Florence,
Via della Lastruccia 3,
50019 Sesto Fiorentino, Italy

Introduction

S100 proteins represent the largest subgroup in the family of calcium-binding proteins bearing EF-hand motifs. A functional EF-hand motif consists of a calcium(II)-binding loop (usually of about 12 amino acids) flanked by two α-helices. S100 proteins contain two EF-hand motifs, one in the N-terminal domain (composed of helix I, loop I, and helix II) and one in the C-terminal domain (composed of helix III, loop II, and helix IV). The two domains are connected by a linker, called a "hinge loop." The first N-terminal EF-hand is unconventional, because its loop is usually composed of 14 amino acids; the second one, in the C-terminal domain, is canonical. A consequence of the longer loop in the N-terminal EF-hand is the different affinity for calcium(II) with respect to the C-terminal EF-hand, due to the different ion coordination. The canonical C-terminal domain in fact binds the ion in a manner similar to calmodulin and troponin-C, resulting in a high calcium affinity [1, 2]. The N-terminal domain mostly binds the ion through main-chain carbonyl groups, in addition to the bidentate side chain of glutamate at the end of the loop, and this reduces the binding affinity up to 100 times [3].

Published online: 03 November 2010

With the exception of calbindin D_{9k}, also known as S100G, which is monomeric, all the other structures of the S100 proteins revealed a homo- and, in some cases, heterodimerization. Some members of the family also form tetramers or larger oligomers. In homodimers, the two subunits are related by a twofold axis of rotation and the major contributors to the dimer interface are helices I and IV of each subunit that are ordered in a X-type four-helix bundle. This relationship is maintained both in the apo state and in the calcium-bound state.

Upon calcium(II) binding most S100 proteins experience a conformational change that mostly involves helix III, which is antiparallel to helix IV in the apo state and rearranges itself to become almost perpendicular in the calcium(II)-bound state. This movement "opens" the structure and exposes a wide hydrophobic cleft that acts as a binding site for targets [4]. Calcium binding to the N-terminal EF-hand, instead, causes only minor alterations of its backbone conformation. On the other hand, calbindin D_{9k} does not undergo changes in its conformation upon calcium(II) binding; S100A7 does not bind calcium in the N-terminal EF-hand [5], as a consequence of the lack of the glutamate residue in the last position of loop I, the carboxylate group of which is essential for coordination of the calcium ion; and S100A10 does not bind calcium in either the N-terminal and or the C-terminal domain. Furthermore, the affinity for calcium in S100A3 is so low ($K_d = 20$ mM) that calcium binding is actually prevented in vivo.

Besides calcium(II), some S100 proteins (S100B [6], S100A2 [7], S100A7 [8], S100A12 [9]) have been shown to bind zinc(II). However, binding of zinc(II) in the cytoplasm is rather unlikely, because of its subnanomolar intracellular concentration. On the other hand, several S100 proteins have been also found in the extracellular space, where the zinc(II) concentration can be much higher [10]; in this respect, zinc was actually reported to modulate the interaction of S100B with the tau protein [11].

S100A16 is the S100 protein most widely distributed in humans, and is highly conserved in mammals [12]. Expression of most S100 proteins is actually highly tissue and cell specific, whereas S100A16 expression has been reported in a wide spectrum of human tissues (including brain), analogously to S100A2, S100A13, and S100A14. Upregulation of S100A16 was found in several cancer tissues, suggesting a function related to malignant transformation or tumor development [12]. S100A16 expression was upregulated in tumors of bladder, lung, thyroid gland, pancreas, and ovary. Furthermore, investigation of S100A16 intracellular localization in human glioblastoma cells revealed an accumulation of the protein within nucleoli and a translocation to the cytoplasm in response to calcium stimulation [13].

Among the S100 family, S100A16 is a "particular" member since it has uncommon characteristics. The N-terminal EF-hand was predicted to be functionally inactive since it comprises 15 amino acids, and lacks the conserved glutamate residue at the last position, analogously to S100A7. The inability of the N-terminal EF-hand to bind calcium was indicated by flow dialysis experiments carried out by Sturchler et al. [13]. Such experiments (performed in a high ionic strength buffer) revealed one Ca^{2+} binding site per subunit, with K_d of 430 μM, which at physiological conditions would be two- to threefold lower, thus becoming very similar to that of many other S100 proteins. Tryptophan fluorescence variations indicated the occurrence of conformational changes upon calcium(II) binding in the C-terminal EF-hand, which lead to the formation of a hydrophobic patch that could involve the hydrophobic residues in helices III and IV and in calcium-binding loop II. They also showed that S100A16 binds zinc(II) in a different site with respect to calcium(II).

Of the 22 members found in the human genome, 17 S100 proteins have genes located in the S100A cluster on chromosome 1q21. Exceptions are S100P (located on chromosome 4p16), S100Z (5q14), S100B (21q22), and calbindin D_{9k} (Xp22) [14]. The human chromosomal region 1q21 is structurally conserved during evolution and exhibits several rearrangements which occur during tumor development. Together with the finding of upregulation of this protein in several cancer tissues [12], this indicates that S100A16 may have a role in the molecular origin of certain types of tumors and thus that it deserves structural and functional characterization studies.

Considering the uncommon behavior of S100A16 with respect to calcium binding, although several S100 protein structures are already available, the structural characterization of human S100A16 in solution has been performed in both the apo and the calcium(II) states. The apo state structure has been also solved in the crystal state. Mobility studies through relaxation rate analysis were also performed in solution. This information represents the starting point for future investigations on the binding with possible targets.

Materials and methods

Protein expression

The gene coding for human S100A16 was generated from complementary DNA using two sets of primers, in two successive runs of polymerase chain reaction (PCR), the second set intended to amplify the specific target sequence within the first, longer, run product. The first set of external primers had the following forward and reverse sequences:

Scheme 1 Amino acid sequence of S100A16. The residues involved in calcium(II) coordination are highlighted

OS116F1 (TGCTGGAGAGGAGGCAGA) and OS116R1 (GGAAGGTCTGGAGGGAGAAG). The second set of specific primers had the following forward and reverse sequences: OS116F2 (AAACATATGTCAGACTGCTAC ACG) and OS116R2 (ATAGAATTCACTAGCTGC TGCTCT). The DNA amplified by PCR was cut with restriction enzymes *Nde*I and *Eco*RI, purified from agarose gel and cloned into plasmid pET21a(+) (Novagen), prepared with the same restriction enzymes. With this expression strategy, the product of the cloned gene has the wild-type sequence of the S100A16 protein (see Scheme 1), without a tag and any additional amino acid.

Vector pET21a(+), containing the human S100A16 gene and cloned to produce the protein without a tag, was transformed in BL21-Gold *Escherichia coli* strain (Novagen). Cells were grown in Luria–Bertani medium at 37°C until an optical density of 0.7 was reached at 600 nm. The protein expression was then induced by adding 1 mM isopropyl β-D-thiogalactopyranoside. The culture was allowed to grow for 4 h and then cells were harvested by centrifugation. The cell pellet was resuspended in lysis buffer [50 mM tris(hydroxymethyl)aminomethane (Tris) pH 8.0, 200 mM KCl, 1 mM dithiothreitol (DTT), 0.5 mM Pefabloc, 10 mM EDTA], and soluble proteins were extracted by sonication followed by centrifugation. The cleared lysate was then precipitated by slowly adding streptomycin sulfate to 1% and centrifugation at 15,000*g* for 20 min. The supernatant was dialyzed in 50 mM Tris pH 7.0, 50 mM KCl, 1 mM DTT, 10 mM EDTA (buffer A) and loaded on a Q Sepharose FF column (Amersham) equilibrated in buffer A and eluted with a linear gradient to 50 mM Tris pH 7.0, 1 M KCl, 1 mM DTT, 10 mM EDTA. The fractions containing S100A16 were collected, added to 2 mM $CaCl_2$, and dialyzed against 50 mM Tris pH 7.4, 200 mM KCl, 1 mM DTT, 2 mM $CaCl_2$ (buffer B). The protein was then purified through hydrophobic exchange with a HiPrep phenyl FF column (Amersham) equilibrated in buffer B and eluted with 50 mM Tris pH 7.4, 200 mM KCl, 1 mM DTT, 5 mM EDTA. A final purification step was performed with size-exclusion chromatography on a HiLoad Superdex 75 16/60 column (Amersham) equilibrated with 20 mM 2-morpholinoethanesulfonic acid (MES) pH 5.5, 200 mM KCl, 1 mM DTT, 1 mM Pefabloc. Protein expression and purity were checked at every step by sodium dodecyl sulfate polyacrylamide gel electrophoresis in 17% polyacrylamide after staining of protein bands with Coomassie blue R-250 against protein marker (Novagen).

Samples of ^{15}N- and $^{13}C,^{15}N$-enriched S100A16 protein were produced as described above except for the use of M9 minimal medium containing $(^{15}NH_4)_2SO_4$ and ^{13}C-glucose as the sole nitrogen and carbon sources.

To express the selenomethionine-labeled S100A16 protein, the recombinant expression vector pET21a(+) was transformed into the methionine-auxotrophic *E. coli* B834(DE3). Cells were grown overnight in 150 mL of selenomethionine medium base supplemented with selenomethionine nutrient mix (Molecular Dimensions) and L-methionine (40 mg L^{-1}). After collection by centrifugation, cells were washed twice with water, resuspended in 1.0 mL water, and added to 1.5 L of the above-mentioned medium supplemented with L-selenomethionine (40 mg L^{-1}). Cells were grown and induced as described above. The recombinant selenomethionine-labeled S100A16 protein was purified as for the native protein except that all buffers were degassed and included a reducing reagent to avoid oxidation of selenomethionine, and a chelator to remove traces of metals that could catalyze oxidation. Full incorporation of selenomethionine was confirmed by mass spectrometry (calculated 11,764.2 Da; observed 11,762.05 Da).

Crystallization, data collection, and structure determination

Crystallization trials on apo wild-type S100A16 and its selenomethionine derivative were performed by the sitting drop method from a solution containing 0.2 M potassium citrate and 20% PEG3350 at 20°C. Hexagonal crystals started to grow overnight.

Several diffraction experiments at −173 °C were performed using synchrotron light radiation. Single-wavelength anomalous diffraction measurements were carried out on the selenium edge wavelength (0.976 Å) at beamline XRD-1 at ELETTRA (Trieste, Italy), and the high-resolution monochromatic data collection was performed at beamline BW7A at DESY-EMBL (Hamburg, Germany).

The selenomethionine derivative crystal diffracted to 2.5-Å resolution and the native crystal diffracted to 2.1-Å resolution; the crystals belonged to the hexagonal space group $P6_1$ (see below) with four molecules (i.e., two functional dimers) in the asymmetric unit and a solvent content of about 55%. The data were collected by the rotation method using 0.5° steps. The two datasets were processed using MOSFLM [15] and scaled using SCALA [16, 17] and both showed a percentage of merohedral twinning of about 10%. The statistics are shown in Table 1.

The analysis of the anomalous Patterson map performed with the program SHELXD [18, 19], using the tenfold redundant dataset collected at the selenium edge (0.976 Å), provided the positions of eight selenium atoms corresponding to two methionines per monomer. The preliminary phases obtained (figure of merit 0.25) were then improved by density modification to a figure of merit of 0.75 using a solvent content of 55% with the program autoSHARP [20, 21]. The first chain tracing after phase refinement performed by ARP/wARP [22] was able to trace 180 residues in the electron density map out of 412; the phases so obtained were then merged with the structure factors of the higher-resolution native dataset and fed into a new chain tracing procedure with BUCCANEER [23], which yielded about 350 residues. The remaining residues were then added and all the side chains were placed manually using XtalView [24]. This procedure was applied to both the possible space groups $P6_1$ and $P6_5$. The latter yielded only a small number of residues traced. Therefore, the correct space group was identified as $P6_1$.

Refinement was carried out using REFMAC5 [17, 25] on the native dataset making use of NCS and TLS restraints and taking twinning into account. Between refinement cycles, the model was subjected to manual rebuilding using XtalView [24]. Water molecules were added using the standard procedure within ARP/wARP [22]. The stereochemical quality of the refined model was assessed using the program Procheck [26]. The Ramachandran plot was of good quality with no residues in the disallowed regions.

The coordinates and structure factors were deposited in the Protein Data Bank under accession code 3NXA.

It is worth mentioning that previous attempts to solve the structure by molecular replacement were unsuccessful. This was not due to a low structural homology of the models used as templates, but to the presence of pseudo-symmetry, due to the fact that the noncrystallographic axis relating the two dimers in the asymmetric unit is close to one of the crystallographic axes. An additional problem is caused by the simultaneous presence of 9–10% merohedral twinning with the operator $k, h, -l$. The latter factor also accounts for R_{cryst} and R_{free} values which are higher than might be expected from the data resolution.

Table 1 Data collection and refinement statistics of the single-wavelength anomalous diffraction (*SAD*) and remote datasets

	SAD dataset	Remote dataset
Synchrotron beamline (detector)	XRD-1 at ELETTRA (MarCCD)	BW7A at DESY-EMBL (MarCCD)
λ (Å)	0.976	1.006
Spacegroup	$P6_1$	$P6_1$
Cell dimensions (Å)	$a = b = 155.96$ $c = 37.09$	$a = b = 156.57$ $c = 38.14$
Resolution (Å)	51.0–2.5 (2.64–2.50)	39.1–2.1 (2.21–2.10)
Total reflections	184,676 (17,630)	340,648 (20,115)
Unique reflections	21,191 (2,821)	29,230 (3,660)
Overall completeness (%)	96.0 (87.4)	91.6 (79.4)
Anomalous completeness (%)	87.1 (58.9)	–
R_{sym} (%)[a]	8.8 (42.3)	9.2 (39.9)
R_{pim} (%)[b]	4.4 (23.3)	2.6 (17.9)
R_{anom} (%)[c]	5.9 (22.2)	–
Multiplicity	8.7 (6.2)	11.7 (5.5)
$\langle I/\sigma(I)\rangle$	5.4 (1.8)	6.0 (1.8)
B factor from Wilson plot (Å^2)	41.3	29.5
Phases FOM before density modification	0.25	–
Phases FOM after density modification	0.75	–
Refinement statistics		
Resolution (Å)		39.1–2.1 (2.15–2.10)
Reflections in working set		26,651 (1,666)
Reflections in test set (9%)		2,650 (173)
R_{cryst}/R_{free} (%)		24.7 (32.9)/29.8 (37.8)
Protein atoms		2,994
Water molecules		96
RMSD bonds (Å)		0.07
RMSD angles (deg)		4.4
Average *B* factor (including metals) (Å^2)		52.50
Residues in most favored/additional allowed/generously allowed/disallowed regions (%)		88.7/10.1/1.2/0.0

Numbers in *parentheses* refer to the high-resolution shell

FOM figure of merit, *RMSD* root mean square deviation

[a] $R_{sym} = \sum_h \sum_l |I_{hl} - \rangle I_h \langle| / \sum_{\mathbf{h}} \sum_l \rangle I_h \langle$

[b] $R_{pim} = \sum_h \sum_l \left(\frac{1}{n_h - 1}\right)^{\frac{1}{2}} |I_{hl} - \rangle I_h \langle| / \sum_h \sum_l \rangle I_h \langle$

[c] $R_{anom} = \sum_{hkl} |\langle I(hkl)\rangle - \langle I(-h-k-l)\rangle| / \sum_{hkl} (\langle I(hkl)\rangle + \langle I(-h-k-l)\rangle)$

Isothermal titration calorimetry

Calcium(II) binding to S100A16 was characterized by measuring the heat changes during the titration of $CaCl_2$ into the protein solution using a MicroCal (Northampton,

MA, USA) VP titration calorimeter. S100A16 and $CaCl_2$ solutions were centrifuged and degassed under vacuum conditions and equilibrated at 37 °C before titration. The sample cell contained 0.2 mM S100A16 dissolved in 20 mM MES buffer (pH 5.5) with 200 mM KCl; the reference cell contained water. The solution of 10 mM $CaCl_2$ was prepared in the same buffer used in the cell sample. Upon equilibration, titrations were performed by injecting 7-μL aliquots of 10 mM ligand ($CaCl_2$) into a 0.2 mM solution of S100A16 using the default injection rate with a 300-s interval between each injection to allow the sample to return to the baseline. The resulting titration curves were corrected using the protein-free buffer control.

NMR spectroscopy and solution structure determination

All NMR experiments for assignments were performed at 25 °C with a Bruker 500 MHz spectrometer equipped with a cryoprobe. Apo and calcium(II)-loaded S100A16 samples (0.6 and 0.8 mM, respectively) were ^{13}C,^{15}N-labeled, in 20 mM MES, 200 mM KCl, and 1 mM DTT buffer (pH 5.5), 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 total correlation spectroscopy, HBHA(CBCACO)HN together with ^{13}C nuclear Overhauser effect spectroscopy (NOESY) heteronuclear single quantum coherence (HSQC) and ^{15}N-NOESY HSQC experiments. Proton–proton distance restraints were derived from the analysis of 2D-NOESY, ^{15}N-NOESY-HSQC, and ^{13}C-NOESY-HSQC spectra acquired with a Bruker 900 MHz spectrometer equipped with a cryoprobe. The spectra were processed using TOPSPIN 2.0 and analyzed with CARA [27]. Backbone dihedral angles were obtained from TALOS+ [28] from the chemical shifts of N, HN, H^{α}, C, C^{α}, and C^{β} nuclei. The structures were calculated using the program CYANA-2.1 [29, 30] by imposing the dimer symmetry constraint (noncrystallographic symmetry constraint). The two subunits in the dimeric structure were linked together through a chain of dummy atoms with zero van der Waals radii. 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 radii, which 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 included for this purpose. 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 structures of the CYANA family were then subjected to restrained energy minimization with AMBER 10 [31]. Nuclear Overhauser effect (NOE) and torsion angle restraints were applied with force constants of 50 kcal mol^{-1} $Å^{-2}$ and 32 kcal mol^{-1} rad^{-2}, respectively. The programs PROCHECK-NMR [32] and WHATIF [33] were used to evaluate the quality of the structures.

Calcium(II) titration was performed with a Bruker 600 MHz spectrometer at 25 °C with 356 μM apo-S100A16 sample. ^{1}H–^{15}N HSQC spectra were acquired for different Ca^{2+} concentrations in solution (0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4, and 12.8 mM).

The coordinates of the apo and calcium(II) solution structures were deposited in the Protein Data Bank under accession codes 2L50 and 2L51, respectively.

Zinc(II) titrations were also performed on both apo-S100A16 and calcium-bound S100A16 with the same experimental conditions as for the calcium(II) titration. ^{1}H–^{15}N HSQC spectra were acquired for different Zn^{2+} concentrations in solution (0.1, 0.3, 0.5, 1, 2, 4, and 8 mM).

Heteronuclear relaxation measurements

^{15}N-R_1, R_2, and steady-state heteronuclear ^{1}H–^{15}N NOEs were measured using a 700 MHz spectrometer using standard pulse sequences [34, 35], at 25 °C. 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 program [36]. 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. R_1, R_2, and NOE values were obtained for 91 out of the 102 assigned backbone ^{N}H resonances for both the apo and the calcium forms. Estimates of the reorientation time were then calculated with the model-free approach [37] and S^2 values were calculated with the program TENSOR2 [38]. Theoretical predictions of ^{N}H R_1 and R_2 values for apo-S100A16 and calcium(II)-loaded S100A16 were calculated using HYDRONMR [39].

Results

Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) experiments were performed to investigate the binding of calcium(II) ions. The binding between apo-S100A16 and Ca^{2+} is endothermic and the reaction proceeds with a positive change in enthalpy. The ITC curve obtained, shown in Fig. 1, is hyperbolic. The best-fit analysis performed using the one binding site model yields an apparent dissociation constant

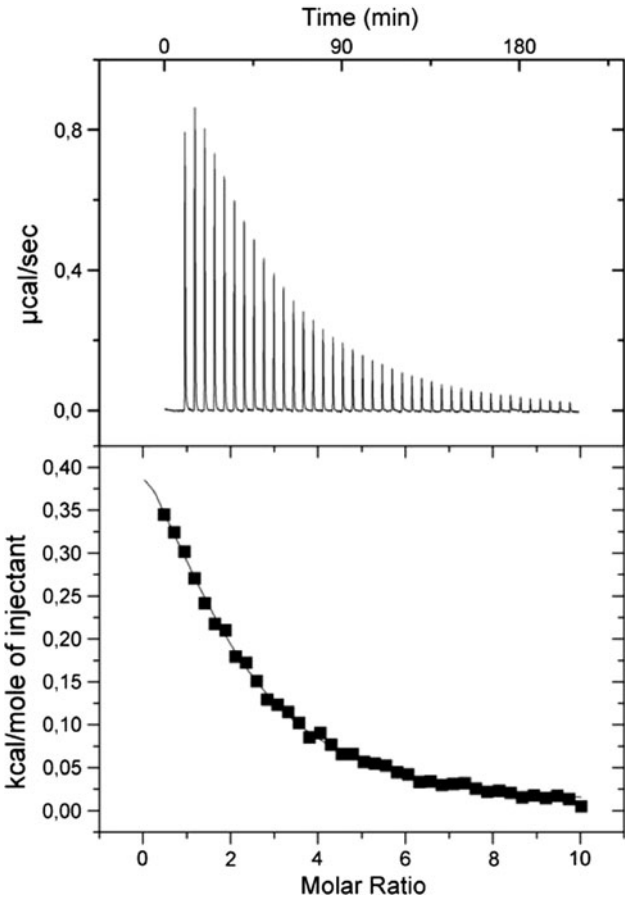


Fig. 1 Isothermograms for the binding of S100A16 to Ca^{2+}. The raw data and the fit to the one binding site model are reported in the *upper panel* and the *bottom panel*, respectively. The fit performed using a sequential two binding sites model is of similar quality. Appropriate background corrections were made to account for the heats of dilution and ionization. All experiments were performed at 25°C

of approximately $(2.7 \pm 0.2) \times 10^{-4}$ M, with 1.88 ± 0.08 binding sites per subunit. The same data were also analyzed with a sequential binding sites model, assuming the presence of two different calcium binding sites per subunit. The analysis with the two binding sites model provided relatively similar ΔH and dissociation constant values, without any significant improvement in the quality of the fit. Therefore, ITC data provide a single binding constant, as previously reported [13] and of similar value, but also suggest the possibility that calcium(II) binding could actually involve both sites in a cooperative way.

NMR resonance assignments

The ^{1}H–^{15}N-HSQC NMR spectra showed well-dispersed signals in both dimensions, which indicated that S100A16 is well folded in both the apo and the calcium-loaded states. All the backbone resonance signals were assigned, except those for Tyr-20 and His-95 in apo-S100A16, Val-23 and Lys-35 in calcium(II)-loaded S100A16, and Ser-2, Lys-32, and Pro-89 in both forms.

Ca^{2+} titration of apo-S100A16

The binding of calcium(II) to apo-S100A16 was monitored by following the changes in the ^{1}H–^{15}N-HSQC NMR spectra of ^{15}N-labeled apo-S100A16 (Figs. S1, S2). The intensity of most peaks in or around both calcium binding regions (from Ser-24 to Ser-34 and from Asp-67 to Glu-78) decreased immediately after the addition of Ca^{2+}, becoming invisible even before reaching a 1:1 ratio between calcium(II) and S100A16. New peaks with increasing intensity then appeared with different chemical shifts when excess Ca^{2+} was added, up to a S100A16-to-calcium(II) ratio of about 1:10. This behavior is indicative of an intermediate exchange regime. In contrast, some other peaks continuously changed their chemical shifts upon increasing the Ca^{2+} concentration up to a 1:10 S100A16-to-calcium(II) ratio, as for systems in the fast exchange regime. These peaks were those experiencing a minor chemical shift perturbation. No peaks showed the typical behavior of the slow exchange regime. The analysis of the chemical shift titration thus indicates that calcium ions perturb several residues in both calcium(II)-binding loops.

Figure S3 shows the change in chemical shift during titration of some fast-exchanging residues, and the corresponding best-fit curves. A dissociation constant of about 3×10^{-4} M can be estimated assuming a cooperative binding model, as found from ITC measurements. This value is in agreement with the value obtained from ITC, and again suggests the presence of two binding sites for calcium(II). Note that both fast-exchanging residues and intermediate-exchanging residues belong to both calcium-binding loops, as the different exchange behavior during the titration depends on the difference in the chemical shift of the apo and calcium forms of the different residues.

Figure 2 shows the chemical shift perturbation on passing from the apo to the calcium(II) form of S100A16. The changes (with an average value of 0.11 ppm) are smaller than for other S100 proteins (average values of, e.g., 0.5 ppm for S100A5 and 0.37 ppm for S100A13). The residues undergoing the largest changes in chemical shifts are located in the two EF-hand loops, the calcium binding sites. The small chemical shift perturbation experienced by residues not belonging to the metal binding sites indicates that the conformational changes occurring between the apo and the calcium-bound forms are smaller than those observed for other S100 proteins.

^{15}N relaxation measurements

The relaxation parameters for apo-S100A16 and calcium-loaded S100A16 are shown in Fig. 3. The reorientation times

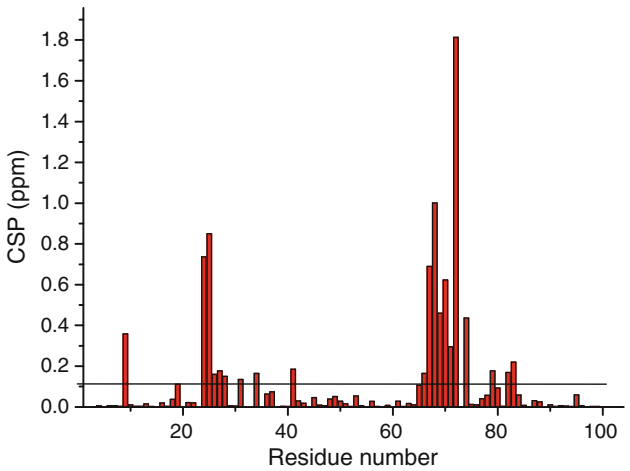


Fig. 2 Chemical shift perturbation (*CPS*) of S100A16 upon calcium(II) binding. The *horizontal line* indicates the average value. Shift perturbations are reported as a weighted average of the amide proton and amide nitrogen shifts using the formula $\Delta\delta = [(\Delta\delta_H)^2 + (\Delta\delta_N/5)^2]^{0.5}$

corresponding to the observed relaxation rates were calculated to be 12.3 ± 1.5 and 12.3 ± 1.8 ns for the apo and calcium-loaded forms of S100A16, respectively, indicating that the protein is dimeric in both forms, and in agreement with the molecular weight and the reorientation times observed for other S100 homodimeric proteins [40–44].

In both apo-S100A16 and calcium-loaded S100A16, the first residues in the N terminus and the residues in the C terminus are poorly structured as a result of their fast internal mobility, revealed by the small or negative NOE values, as well as by the large R_1 and the small R_2 values. Fast motion is also detected for some residues at the beginning of helix II (Ser-37, Phe-38 in the apo form; Ser-36, Phe-38 in the calcium form). Sizable motion is detected for loop L1 of the N-terminal EF-hand motif and linker L2 between the two EF-hand motifs.

Upon calcium binding, several residues are subject to an increase in mobility. Faster internal motions are present in loop L1 (the ^1H–^{15}N-NOE values decrease with respect to the apo form), whereas the residues at the end of helix IV (Gly-84, Ile-86, Ile-90, and Ala-91) and Asp-67 experience motions on a slower timescale, as indicated by the significantly larger R_2 value compared with the average values observed for the other residues. A reduction in mobility is, in contrast, observed upon calcium binding for the residues in loop L3 of the C-terminal EF-hand motif.

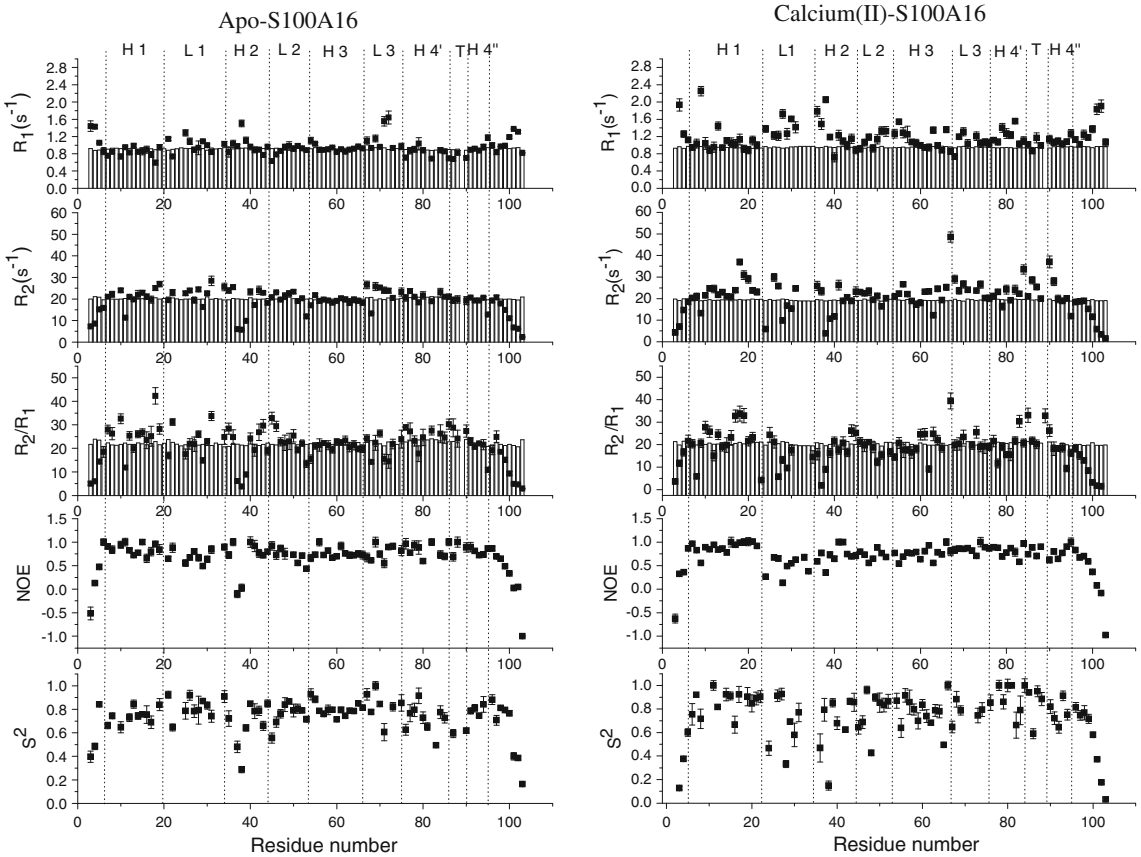


Fig. 3 Sequential plot of the experimental relaxation parameters of apo-S100A16 and calcium(II)-loaded S100A16. The values calculated with HYDRONMR are shown as *bars*

Solution structure of apo-S100A16 and calcium-loaded S100A16

The solution structures of human S100A16 in the apo and calcium-loaded forms were calculated from a total of 1,177 and 1,167 meaningful intrasubunit upper distance limits and 89 and 94 intersubunit upper distance limits for the apo and calcium forms, respectively. Few NOE patterns were detected for the residues in loop L1 between helix I and helix II and at the C terminus, consistent with the observed mobility in these regions. In the calcium form, the Ca^{2+} ions were restrained to be within 3 Å from the oxygen ligand atoms (O of Val-23, Tyr-26, Leu-28, and Lys-32 for the N-terminal Ca^{2+} binding site; OD1 of Asp-67 and Asn-69; OD1 and OD2 of Asp-71; O of Arg-73; OE1 and OE2 of Glu-78 for the C-terminal Ca^{2+} binding site).

Since no unique NOEs were detected for one subunit and not for the other, the calculations were performed by imposing the dimer symmetry constraint into the CYANA calculation. The root mean square deviation (RMSD) from the mean structure for the structured regions of the dimeric protein is 0.8 ± 0.1 Å (backbone) and 1.2 ± 0.1 Å (heavy atoms) for apo-S100A16 (residues 7–23, 35–97 of both subunits) and 0.7 ± 0.2 Å (backbone) and 1.1 ± 0.1 Å (heavy atoms) for calcium(II)-loaded S100A16 (residues 7–23, 35–97 of both subunits). PROCHECK-NMR and WHATIF programs were used to validate the structures on the Web site https://nmr.cmbi.ru.nl/icing/iCing.html. More than 98% of the residues in both apo and calcium(II) structure families were located in the allowed regions of the Ramachandran plot. The statistical analysis is reported in Table 2. The not excellent quality is common to many S100 proteins, probably owing to the property of this class of proteins (and of other signaling proteins based on the EF-hand domain) to change conformation depending on the calcium state. The relaxation rates calculated with HYDRONMR [39] from the minimized mean structures under the assumption of no internal motions, shown in Fig. 3, are in overall agreement with the averaged experimental values. This confirms that the protein is dimeric. On the other hand, the differences between the calculated and observed relaxation rates make it easier to appreciate the presence of mobility for some residues (see "^{15}N relaxation measurements") [45–51].

The calculated families of structures are shown in Fig. 4. In both forms, the four helices of the two EF-hand motifs of each subunit are well defined, whereas loop L1 of the fist EF-hand motif is less well defined. These results are in line with the relaxation results. Helix IV is interrupted by residue Pro-89, after which the helical arrangement starts again.

Crystal structure of apo-S100A16

The crystal structure of apo-S100A16 was solved as described in the "Materials and methods." The statistics are reported in Table 1. The structure generally shows a well-defined electron density map for the four helices of the two EF-hand motifs of each subunit except for residues from 51 to 71 of monomer D, comprising helix III and part of the loop between helix III and IV. This is consistent with a very high degree of mobility of these regions in the crystal lattice, as also indicated by the *B* factors. To obtain reasonably low *R* values, the geometry weight had to be lowered in the refinement procedure, and this resulted in

Table 2 Structural restraints and statistical analysis

	Apo-S100A16	Ca(II)-S100A16
NOE upper distance limits		
Intrasubunit	1,177	1,167
Intraresidue	510	560
Interresidue		
Sequential ($\|i-j\|=1$)	288	288
Medium range ($\|i-j\|<4$)	236	224
Long range ($\|i-j\|>5$)	143	95
Intersubunit	89	94
Dihedral angle restraints per subunit		
φ	64	62
ψ	64	62
Average RMSD from the mean (Å)		
Backbone	1.2 ± 0.2[a]	1.1 ± 0.3[a]
	0.8 ± 0.1[b]	0.7 ± 0.2[b]
Heavy	1.7 ± 0.3[a]	1.7 ± 0.4[a]
	1.2 ± 0.1[b]	1.1 ± 0.1[b]
Residual CYANA target function (Å^2)	0.7 ± 0.1	0.7 ± 0.1
Structure analysis		
Residues in most favorable regions (%)	82.1[a]	81.5[a]
	87.5[b]	87.2[b]
Residues in allowed regions (%)	13.9[a]	14.1[a]
	11.2[b]	11.4[b]
Residues in generously allowed regions (%)	2.2[a]	2.7[a]
	0.7[b]	0.9[b]
Residues in disallowed regions (%)	1.8[a]	1.7[a]
	0.6[b]	0.5[b]
Structure *Z* scores		
2nd-generation packing quality	−2.6 ± 0.3	−2.5 ± 0.4
Ramachandran plot appearance	−4.7 ± 0.5	−4.4 ± 0.4
χ_1/χ_2 rotamer normality	−5.7 ± 0.3	−5.3 ± 0.4
Backbone conformation	−0.8 ± 0.5	−0.7 ± 0.5
RMS *Z* scores		
Bond length	1.187 ± 0.002	1.184 ± 0.003
Bond angles	0.83 ± 0.01	0.86 ± 0.02
Omega angle restraints	1.9 ± 0.1	2.0 ± 0.2
Side chain planarity	2.1 ± 0.3	2.0 ± 0.2
Improper dihedral distribution	1.27 ± 0.05	1.33 ± 0.05
Inside/outside distribution	1.04 ± 0.03	1.01 ± 0.01

NOE nuclear Overhasuer effect, *RMS* root mean square

[a] Values were calculated in the sequence range 7–95 of both subunits

[b] Values were calculated in the sequence ranges 7–23 and 35–95 of both subunits

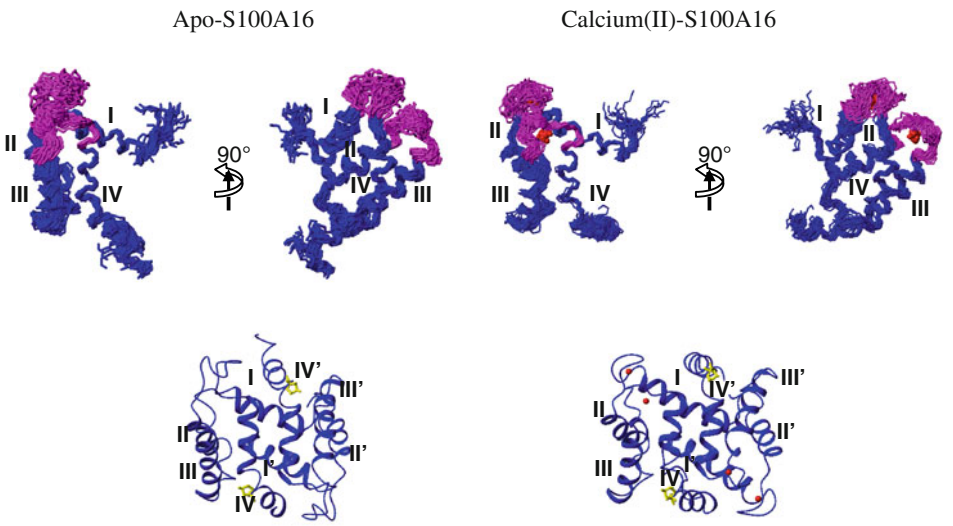


Fig. 4 Solution structural families of the S100A16 subunit and ribbon representation of the homodimer mean structures of the protein in the apo and calcium(II)-bound forms. *Roman numerals* indicate the helix numbers. Calcium ligand loops are shown in *magenta*. Pro-89 is shown in *yellow*

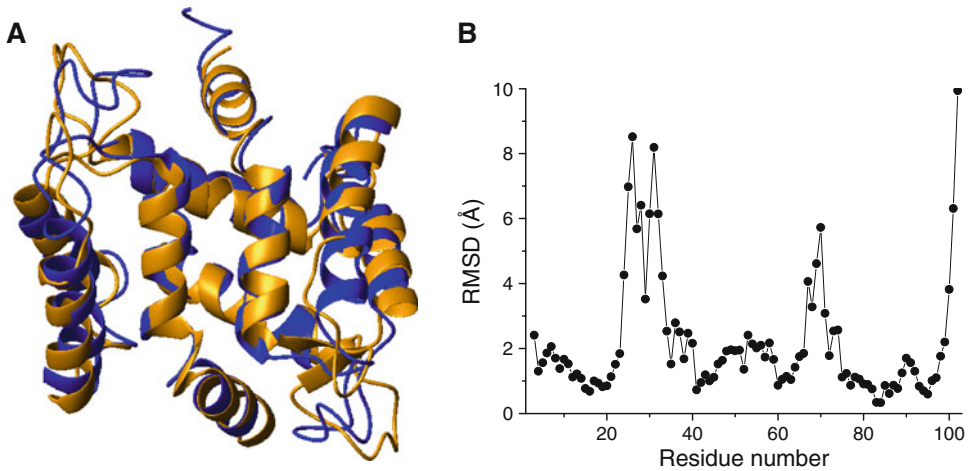


Fig. 5 **a** Three-dimensional structure of one of the two homodimers in the crystal structure (in *yellow*) superimposed on the mean solution NMR structure (in *blue*) of apo-S100A16, and **b** backbone root mean square deviation (*RMSD*) between the crystal structure and the mean average structure

rather high RMSD for bond lengths and angles. Twinning (mainly) and pseudosymmetry (partially) are likely the reasons for this deviation from standard statistics for a crystal structure of analogous resolution.

The superposition between the mean NMR apo structure and the X-ray structure shows that the solution and solid-state structures of apo-S100A16 are in overall agreement, with the exception of loop L1 and loop L3 regions, as shown in Fig. 5. The mean backbone RMSD between the two structures of each subunit is 2.6 Å in the whole range of protein residues, but if we consider only the sequence ranges 7–23 and 35–95, it decreases to 1.7 Å, and if we exclude the two above-mentioned regions (residues 24–34 and 66–73), besides the very first and last residues at the N terminus and C terminus, which are intrinsically mobile, the RMSD decreases to 1.3 Å, indicating that the structures are in good agreement (Fig. 5). Furthermore, the disagreement is mainly due to local discrepancies rather than to overall changes in the interhelical angles (see Table 3).

Zn^{2+} and Cu^{2+} titration of apo-S100A16

After addition of Zn^{2+} to apo-S100A16, the peak intensity of the residues located in the hinge loop, in the turn region of the last helix, and at the N terminus started decreasing appreciably at a S100A16-to-zinc(II) ratio of 1:1, and some peaks disappeared when a 1:3 ratio was reached. No new peaks appeared during the whole titration, and all other peaks remained unperturbed. Similar changes were observed during the zinc(II) titration of calcium(II)-loaded S100A16.

S100A16 should thus bind zinc(II) with low affinity [dissociation constant greater than 10^{-4} M for the apo form and even larger for the calcium(II) form]. Some residues in the hinge loop (His-48) and at the N terminus (Cys-4, Glu-9) of the other subunit may constitute the Zn^{2+} ligands.

Copper(II) titration of apo-S100A16 was also attempted but the protein immediately precipitated after addition of copper(II).

Table 3 Angles between different helices, the directions of which are defined by the eight residues immediately preceding or following each EF-hand loop, calculated from the mean solution NMR structure (the errors are calculated from the standard deviations within the 30 structures of the families)

	Apo-S100A16 (deg)	Ca(II)-S100A16 (deg)
I–II	136 ± 3 (128 ± 2)	142 ± 4
I–III	56 ± 4 (64 ± 3)	59 ± 4
I–IV	118 ± 3 (116 ± 1)	114 ± 4
II–III	157 ± 5 (163 ± 2)	144 ± 4
II–IV	52 ± 6 (37 ± 1)	60 ± 6
III–IV	148 ± 3 (153 ± 1)	150 ± 4
I–I′	136 ± 3 (154 ± 1)	138 ± 6
IV–IV′	156 ± 4 (158 ± 2)	166 ± 4

The values in *parentheses* refer to the angles calculated from the X-ray structure

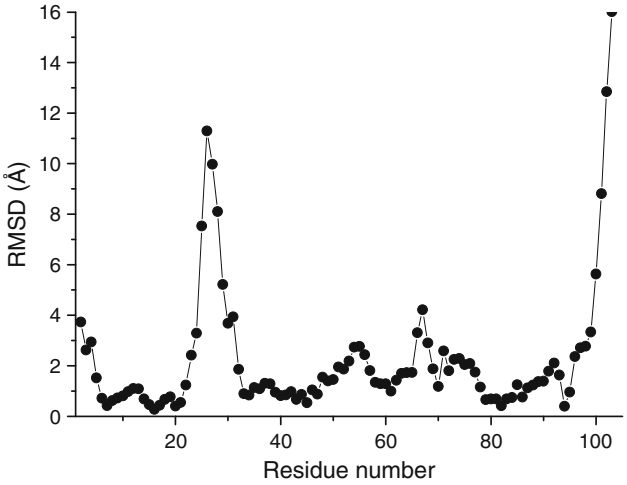


Fig. 6 Backbone RMSD between the apo-S100A16 and the calcium(II)-loaded S100A16 mean solution structures

Discussion

In both apo-S100A16 and calcium-loaded S100A16, dimerization mostly occurs through interactions between helices I, I′, IV, and IV′, which form an X-type helix bundle. Hydrophobic residues Trp-80 and Ile-83 in helix IV make several contacts with Leu-8, Val-12, and Leu-15 in helix I′ and with Trp-80 and Ile-83 in helix IV′ of the other subunit. Residues Glu-45, Leu-46, His-48, and Met-49 in the hinge loop between helices II and III also make contacts with residues near the N terminus of helix I′ of the other subunit. In the S100A16 dimer, all these interactions align helices I and IV in opposite directions to helices I′ and IV′, respectively.

The overall fold of the protein in the apo form is in agreement with the previously known structures for other S100 proteins [41, 44, 49, 52–54]. However, and differently from most of the other S100 proteins, it is apparent that after calcium binding S100A16 does not undergo any major conformational changes. Indeed, the backbone RMSD between the apo and the calcium(II)-loaded solution structures in the structured regions of the dimer (7–23, 35–95 of both subunits) is only 1.6 Å (Fig. 6). The C-terminal EF-hand motif does not move to the open conformation upon calcium(II) binding as shown experimentally, for instance, by the presence of strong NOEs between Ala-59 in the third helix and Ile-86 in the fourth helix.

The largest change in the solution structure of S100A16 upon calcium binding is in the angle between helices II and III, which varies from 157 ± 5° in the apo form (163 ± 2° in the crystal structure) to 144 ± 4° in the calcium-loaded form (see Table 3). The angles are measured by defining the directions of the α-helices in each EF-hand motif from the eight residues immediately preceding and following each EF-hand loop [55]. For solution structures, such values are calculated from the mean NMR structure and the corresponding errors from the standard deviation observed within the structures of the families. The approximately 15–20° difference in the angle between helices II and III upon calcium coordination is significantly smaller than that measured for S100A13 (40°), which is the closest neighbor of S100A16 in the phylogenetic tree.

The angle between helices III and IV is 148 ± 3° in the apoprotein (153 ± 1° if measured in the apo crystal structure), as expected for the almost antiparallel arrangement typical of EF-hand motifs in the absence of calcium. In other S100 proteins, such as S100A3, S100A5, and S100A13, such an angle typically changes by 30–50° upon calcium binding [44, 49], so the two helices become almost perpendicular [53, 56–58]. In contrast, in the calcium-loaded S100A16, the angle between helices III and IV is 150 ± 4°, so they remain almost antiparallel. Correspondingly, helices I and I′ and helices IV and IV′ make similar angles in both the apo and the calcium forms, differently from most S100 proteins.

As shown in Fig. 6, there is a significant conformational difference at the C terminus between the mean solution structures of S100A16 in the apo and calcium forms. This difference is due to the large mobility in solution of the residues after the last helix.

As already seen, the differences between the X-ray structure and the NMR structure of apo-S100A16 are mainly in the loops and in the N-terminal and C-terminal regions, due to disorder of these protein regions in solution, in this case likely due to mobility. The global orientation of the helices is, in contrast, very similar, as shown in Table 3. The global orientation of the helices is actually the main criterion to judge how much conformational change takes places.

The superposition of the NMR structure onto the crystal structure and the following symmetry expansion (coherently

with the crystallographic symmetry) does not show any significant difference in the crystal packing contacts with respect to those present in the crystal structure. This implies that the structural differences in the above-mentioned regions in the solid state are not due to packing contacts but are related to an intrinsic mobility of those regions.

A principal component analysis of the six interhelical angles representing the reciprocal orientation of the four helices [55] clearly shows the peculiar features of S100A16 within the EF-hand family. With use of the first two principal components, it is found that EF-hand proteins are clearly clustered into two subgroups (closed and open) which are characterized by the protein metal content, i.e., the apo and calcium-loaded forms. The principal component values for the two forms of S100A16 in solution were calculated from the interhelical angles reported in Table 3 and plotted together with the values previously calculated for all the other S100 proteins [49] (Fig. 7), by using the same coefficients for the interhelical angles reported in Babini et al. [55]. The principal component plot shows that apo-S100A16 is regularly positioned with respect to the other apo S100 proteins, whereas in the calcium-loaded form it is still located in the subgroup corresponding to the closed structures in the apo state. Therefore, and at variance with all the other S100 proteins, the calcium-loaded form maintains a similar overall arrangement as the apo form. It is to be noted that the only other S100 proteins not regularly placed in the principal component plot are calbindin D_{9k} and S100A10. However, and at variance with S100A16, for both of them the apo form maintains an arrangement similar to that of the calcium-loaded form. In other words, calbindin D_{9k} and S100A10 are already in the open conformation even in the absence of calcium, whereas, in contrast, S100A16 is the first example of a calcium-loaded form which remains almost as closed as the apo form.

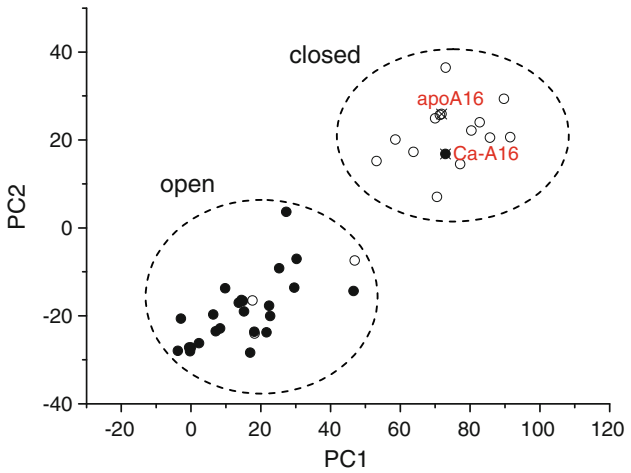


Fig. 7 Principal component plot for the S100 proteins derived from principal component analysis of the six interhelical angles. Apoproteins (S100A1, S100A2, S100A3, S100A4, S100A5, S100A6, S100A10, S100A11, S100A13, S100A16, S100B, calbindin D_{9k}) are indicated with *open circles* and calcium(II)-bound proteins (S100A1, S100A4, S100A5, S100A6, S100A7, S100A8, S100A9, S100A12, S100A13, S100A16, S100P, S100B, calbindin D_{9k}) are indicated with *solid circles*. The two *open symbols* not regularly placed with respect to the other correspond to calbindin D_{9k} and S100A10 in the apo form. The *solid symbol* not regularly placed with respect to the other corresponds to S100A16 in the calcium(II)-bound form. The data are based on the structural information reported in Table 3 and on data reported in Bertini et al. [49]. *PC1* first principal component, *PC2* second principal component

In most S100 proteins the two calcium binding sites are the classic EF-hand C-domain binding site and the S100-specific N-domain binding site. The former contains highly conserved calcium ligand residues at positions 1, 3, 5, 7, and 12, and has a larger affinity for the metal. The latter is a 14-residue motif where the calcium ligands are the backbone oxygen atoms of the residues at positions 1, 4, 6, and 9 and, in most cases, two side chain oxygen atoms of the residue at position 14 (usually Glu). The N-domain binding site of S100A16 lacks the glutamate at this last position (see Scheme 1). This is expected to sizably decrease the calcium binding affinity, because two important ligands are missing. Furthermore, the N-terminal EF-hand comprises 15 amino acids instead of 14, owing to the insertion, unique for S100A16, of residue Leu-28, and the ligand at position 9 is replaced by a ligand at position 10. S100A16 has been reported to bind one calcium(II) ion only for each subunit, i.e., that in the C-terminal EF-hand, through flow dialysis experiments (buffer 50 mM Tris–HCl, pH 7.5, 500 mM KCl) [13]. The present study suggests that in our conditions S100A16 indeed retains the ability to bind a calcium ion (with low affinity) also in the N-terminal EF-hand motif even without the glutamate at position 14. The calcium titration followed by NMR spectroscopy indicates that most of the residues on both calcium binding sites are in an intermediate or fast exchange regime. Chemical shifts changed until 10 equiv of calcium(II) per subunit was added, pointing out the low binding affinity for both sites.

The present observations allow us to make some general comments on the energetics involved in the calcium-triggered conformational changes that characterize the functional role of S100 proteins. To do so, reference can be made to Fig. 8a, where the calcium binding and the conformational changes are separated. As illustrated in the figure, the equilibrium constant K for the apo closed form and the calcium(II) open form is the product of the equilibrium constant for the apo and the calcium forms in the closed state (K_1) multiplied by that for the closed and open forms in the calcium(II)-bound state (K_3): $K = K_1 K_3$ ($=K_2K_4$). For "normal" S100 proteins, $K_2 < 1$ (i.e., the apo closed form is more stable, see Fig. 8b) and $K_3 \gg 1$ (i.e.,

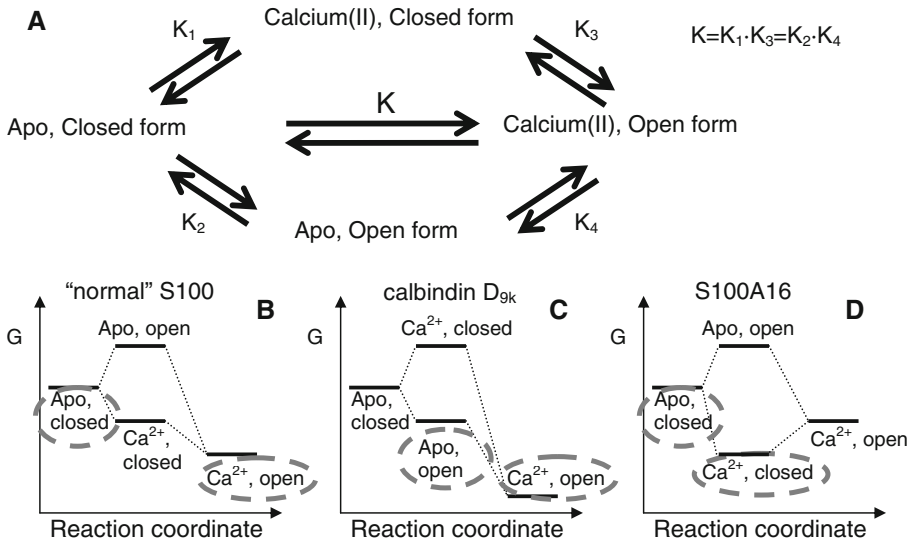


Fig. 8 Equilibrium constants (**a**) and energy levels (**b**–**d**) for the of open and closed forms of the S100 structures in the apo and calcium(II)-bound states. Observable forms of "normal" S100 proteins (**b**), calbindin D_{9k} (**c**), and S100A16 (**d**) are highlighted with *dashed lines*

the calcium open form is more stable). In the case of calbindin D_{9k}, the apoprotein is more stable in the open form ($K_2 > 1$), i.e., in a "calcium-ready" form (Fig. 8c). Therefore, calcium binding is enhanced, as $K_4 \gg 1$. Conversely, S100A10, which also exists as apoprotein in a "calcium-ready" form, has lost its ability to bind calcium. It has been speculated that S100A10 is a structural protein that needs to always be in the open form and does not need to be opened by a signal, and therefore has lost its ability to bind calcium. Indeed, the first putative binding loop lacks three residues and cannot bind Ca^{2+} [59], and some amino acid replacements in the second putative binding loop (Asp-Cys at position 61, Glu-Ser at position 70 with respect to calbindin D_{9k}) hamper the ability of this loop to bind calcium [60]. S100A10 is in fact in a permanently activated state, having hydrophobic residues exposed even in the absence of Ca^{2+} [60, 61], which allow the protein to act as a linker tethering certain transmembrane proteins to annexin A2 and thereby assisting their traffic to the plasma membrane and/or their firm anchorage at certain membrane sites [62]. So, for both calbindin D_{9k} and S100A10, $K_2 > 1$. The case of S100A16 investigated here is an unprecedented case of $K_3 < 1$, i.e., the closed calcium-loaded form is more stable (Fig. 8d). This, of course, implies that $K_1 > 1$, despite the fact that the collective binding of the two calcium ions is relatively weak. In turn, this suggests that $K_1 > 1$ also for the "normal" S100 proteins, and that their higher calcium affinity is due to a favorable combination of both $K_1 > 1$ and $K_3 > 1$. In other words, S100A16 is somehow the opposite of calbindin D_{9k}. Whereas in normal S100 proteins calcium binding is described by the product $K = K_1 K_3$, in the case of calbindin D_{9k} and S100A16 calcium binding is only described by either K_4 or K_1, respectively. The relatively small calcium affinity of S100A16 is thus due to the low value of K_3, which makes the binding only dependent on K_1.

The presence of hydrophobic interactions represents an important factor in moving the equilibrium between the

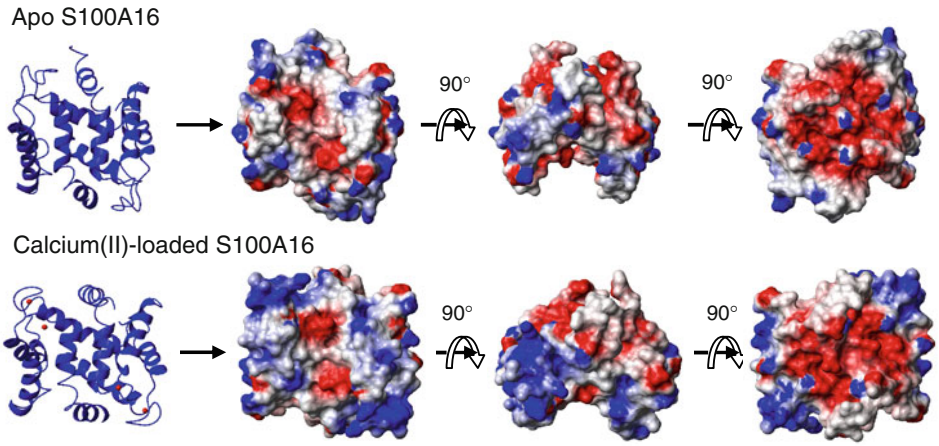


Fig. 9 Electrostatic surface representation of the S100A16 dimer

open and the closed forms in EF-hand motifs. In S100 proteins this equilibrium depends mainly on the presence/absence of interactions between the hydrophobic residues of the third and fourth helices. In S100A16 the number of hydrophobic residues present in the third helix is larger than for other S100 proteins. In the closed form of S100A16, strong interactions among hydrophobic residues are actually present between the third helix (residues Ala-58, Ala-59, Leu-62, Ile-63,and Leu-66) and the fourth helix (residues Leu-82 and Ile-86). These interactions are likely to make the closed structure of the second EF-hand particularly stable, so even upon calcium(II) binding such a conformation is not disrupted.

In S100A16, helix IV has the same length in both the apo and the calcium-bound states, differently from some other S100 proteins (S100A5, S100A6, and S100B), where it is longer in the calcium(II)-bound form than in apo form [49, 53, 63, 64]. The helix is interrupted and divided into two short helices by an 84-89 (Gly-Gly-Ile-Thr-Gly-Pro) sequence motif with three glycine residues and one proline residue. In water-soluble proteins, proline is a potent helix breaker [65]. It either breaks or kinks a helix because it cannot donate an amide hydrogen bond, and because its side chain sterically interferes with the backbone of the preceding turn. This forces a bend of about 30° in the helix axis [66, 67]. Furthermore, the glycine residues also tend to disrupt helices because their high conformational flexibility makes it entropically expensive to adopt the relatively constrained α-helical structure and because they lack hydrophobic stabilization [68].

Upon calcium binding, the global shape of the dimeric protein changes, as a result of the structural differences, as well as of the change in the distribution of surface charges. The electrostatic potential surface calculation, the results of which are shown in Fig. 9, was performed with MOLMOL [69] after inclusion of the calcium(II) charge into the AtomCharge setup file. Red and blue areas indicate negatively and positively charged regions, respectively. On passing from the apo to the calcium-loaded form, hydrophobic and positively charged residues are more exposed, whereas negatively charged residues are somewhat less exposed. These features may be important for the binding capability of the protein in the two forms. In fact, each S100 protein seems to show a peculiar surface charge and hydrophobic distribution as well as different changes upon calcium binding, ranging from exposing a more hydrophobic surface, to a larger negatively charged surface, or to a different position of charged and hydrophobic residues on the surface. This diversity is likely to be linked to the their different target specificities.

In conclusion, we have shown that the homodimeric structure of human S100A16 is subject to conformational rearrangements upon calcium(II) binding that are much smaller than those observed for most of the other S100 proteins. This is likely to be related to the weak binding affinity of the protein for the calcium(II) ions, and to the fact that the closed structure of the second EF-hand is particularly stable in the presence of strong hydrophobic interactions, so even upon calcium(II) binding such conformation is not disrupted.

Acknowledgments This work was supported by MIUR-FIRB contracts RBLA032ZM7 and RBIP06LSS2, and by European Commission contracts EU-NMR 026145 and SPINE2-COMPLEXES 031220.

References

1. Santamaria-Kisiel L, Rintala-Dempsey AC, Shaw GS (2006) Biochem J 396:201–214
2. Donato R (2003) Microsc Res Tech 60:540–551
3. Donato R (1986) Cell Calcium 7:123–145
4. Nelson MR, Chazin WJ (1998) Biometals 11:297–318
5. Brodersen DE, Etzerodt M, Madsen P, Celis JE, Thøgersen HC, Nyborg J, Kjeldgaard M (1998) Structure 6:477–489
6. Wilder PT, Baldisseri DM, Udan R, Vallely KM, Weber DJ (2003) Biochemistry 42:13410–13421
7. Randazzo A, Acklin C, Schafer BW, Heizmann CW, Chazin WJ (2001) Biochem Biophys Res Commun 288:462–467
8. Brodersen DE, Nyborg J, Kjeldgaard M (1999) Biochemistry 38:1695–1704
9. Moroz OV, Burkitt W, Wittkowski H, He W, Ianoul A, Novitskaya V, Xie J, Polyakova O, Lednev IK, Shekhtman A, Derrick PJ, Bjoerk P, Foell D, Bronstein IB (2009) BMC Biochem 10:11
10. Hwang JJ, Park MH, Choi SY, Koh JY (2005) J Biol Chem 280:11995–12001
11. Yu WH, Fraser PE (2001) J Neurosci 21:2240–2246
12. Marenholz I, Heizmann CW (2004) Biochem Biophys Res Commun 313:237–244
13. Sturchler E, Cox JA, Durussel I, Weibel M, Heizmann CW (2006) J Biol Chem 281:38905–38917
14. Ridinger K, Ilg EC, Niggli FK, Heizmann CW, Schäfer BW (1998) Biochim Biophys Acta 1448:254–263
15. Leslie AGW (1991) In: Moras D, Podjarny AD, Thierry J-C (eds) Molecular data processing. Oxford University Press, Oxford
16. Evans PR (1993) Proceedings of the CCP4 study weekend. In: Sawyer L, Isaacs N, Bailey S (eds) Data collection and processing, pp 114–122
17. Collaborative Computational Project N (1994) Acta Crystallogr D50:760–763
18. Schneider TR, Sheldrick GM (2002) Acta Crystallogr D 58:1772–1779
19. Sheldrick GM (2008) Acta Crystallogr A 64:112–122
20. Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Methods Mol Biol 364:215–230
21. Bricogne G, Vonrhein C, Flensburg C, Schiltz M, Paciorek W (2003) Acta Crystallogr D 59:2023–2030
22. Perrakis A, Morris RJH, Lamzin VS (1999) Nat Struct Biol 6:458–463
23. Cowtan K (2006) Acta Crystallogr D 62:1002–1011
24. McRee DE (1999) J Struct Biol 125:156–165
25. Murshudov GN, Vagin AA, Dodson EJ (1997) Acta Crystallogr D53:240–255
26. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) J Appl Crystallogr 26:283–291

27. Keller R (2004) The computer aided resonance assignment tutorial. CANTINA, Goldau
28. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) J Biomol NMR 44:213–223
29. Guntert P (2004) Methods Mol Biol 278:353–378
30. Herrmann T, Güntert P, Wüthrich K (2002) J Mol Biol 319:209–227
31. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister CE, Caldwell JW, Ross WS, Kollman PA (2008) AMBER 10. University of California, San Francisco
32. Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM (1996) J Biomol NMR 8:477–486
33. Vriend G (1990) J Mol Graphics 8:52–56
34. Kay LE, Torchia DA, Bax A (1989) Biochemistry 28:8972–8979
35. Barbato G, Ikura M, Kay LE, Pastor RW, Bax A (1992) Biochemistry 31:5269–5278
36. Goddard TD, Kneller DG (2000) SPARKY 3. University of California, San Francisco
37. Lipari G, Szabo A (1982) J Am Chem Soc 104:4546–4559
38. Dosset P, Hus JC, Marion D, Blackledge M (2001) J Biomol NMR 20:223–231
39. Garcia de la Torre JG, Huertas ML, Carrasco B (2000) J Magn Reson 147:138–146
40. Inman KG, Baldisseri DM, Miller KE, Weber DJ (2001) Biochemistry 40:3439–3448
41. Zhukov I, Ejchart A, Bierzynski A (2008) Biochemistry 47:640–650
42. Dutta K, Cox CJ, Basavappa R, Pascal SM (2008) Biochemistry 47:7637–7647
43. Bertini I, Fragai M, Luchinat C, Parigi G (2000) Magn Reson Chem 38:543–550
44. Arnesano F, Banci L, Bertini I, Fantoni A, Tenori L, Viezzoli MS (2005) Angew Chem Int Ed 44:6341–6344
45. Bertini I, Fragai M, Luchinat C, Melikian M, Mylonas E, Sarti N, Svergun D (2009) J Biol Chem 284:12821–12828
46. Bertini I, Calderone V, Fragai M, Jaiswal R, Luchinat C, Melikian M, Mylonas E, Svergun D (2008) J Am Chem Soc 130:7011–7021
47. Bertini I, Gupta YK, Luchinat C, Parigi G, Schlörb C, Schwalbe H (2005) Angew Chem Int Ed 44:2223–2225
48. Luchinat C, Parigi G (2007) J Am Chem Soc 129:1055–1064
49. Bertini I, Dasgupta S, Hu X, Karavelas T, Luchinat C, Parigi G, Yuan J (2009) J Biol Inorg Chem 14:1097–1107
50. Brüschweiler R (2003) Curr Opin Struct Biol 13:175–183
51. Bernadò P, Garcìa de la Torre J, Pons M (2002) J Biomol NMR 23:139–150
52. Smith SP, Shaw GS (1998) Structure 6:211–222
53. Otterbein L, Kordowska J, Witte-Hoffmann C, Wang CL, Dominguez R (2002) Structure 10:557–567
54. Drohat AC, Baldisseri DM, Rustandi RR, Weber DJ (1998) Biochemistry 37:2729–2740
55. Babini E, Bertini I, Capozzi F, Luchinat C, Quattrone A, Turano M (2005) J Proteome Res 4:1961–1971
56. Marenholz I, Heizmann CW, Fritz G (2004) Biochem Biophys Res Commun 322:1111–1122
57. Maler L, Sastry M, Chazin WJ (2002) J Mol Biol 317:279–290
58. Bhattacharya S, Chazin WJ (2003) Structure 11:738–739
59. Gerke V, Weber K (1985) EMBO J 4:2917–2920
60. Rety S, Sopkova J, Renouard M, Osterloh D, Gerke V, Tabaries S, Russo-Marie F, Lewit-Bentley A (1999) Nat Struct Biol 6:89–95
61. Kube E, Becker T, Weber K, Gerke V (1992) J Biol Chem 267:14175–14182
62. Rescher U, Gerke V (2008) Pflugers Arch J Physiol 455:575–582
63. Smith SP, Shaw GS (1998) Structure 6:211–222
64. Kilby PM, Van Eldik LJ, Roberts GC (1996) Structure 4: 1041–1052
65. Jacob J, Duclohier H, Cafiso DS (1999) Biophys J 76:1367–1376
66. Richardson JS (1981) Adv Protein Chem 34:167–339
67. MacArthur MW, Thornton JM (1991) J Mol Biol 218:397–412
68. Blaber M, Zhang XJ, Matthews BW (1993) Science 260:1637–1640
69. Koradi R, Billeter M, Wüthrich K (1996) J Mol Graphics 14:51–55

3.3

IS HUMAN MULTIPROTEIN BRIDGING FACTOR 1 A CALMODULIN TARGET? IN VITRO DETECTION BY NMR SPECTROSCOPY AND CAM-AGAROSE AFFINITY CHROMATOGRAPHY

Elena Babini,[1]* Francesco Capozzi,[1,2] Xiaoyu Hu,[2] Giacomo Parigi,[2] Massimiliano Vignali[1]

[1]Department of Food Science, University of Bologna, Piazza Goidanich 60, 47521, Cesena, Italy.

*e-mail: elena.babini2@unibo.it

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

ABSTRACT

The human Multiprotein Bridging Factor 1 (hMBF1) has been established in different cellular types to have the role of transcriptional coactivator. It is also a putative Calmodulin (CaM) target, reported to bind CaM in its calcium free state, but little is known about the structural features and the biological relevance of this interaction. We applied NMR to investigate the interaction between the two proteins in solution and compared the results with those obtained with CaM-agarose affinity chromatography. No changes in ^{1}H-^{15}N HSQC spectrum of both apo-CaM and Ca^{2+}-CaM upon addition of hMBF1, prove that the two proteins do not interact in vitro. These results were confirmed by CaM-agarose affinity chromatography when operating under the same conditions. The discrepancy with respect to present and previous experiments performed with CaM-agarose affinity chromatography depends on different experimental parameters suggesting that particular attention must be paid when CaM, or other immobilized proteins, are used to measure their affinity with putative partners. These results also implies that if an interaction between the two proteins exists in vivo, as reported for hMBF1 of endothelial cells, it might involve a posttranslational modified form of the proteins or it relies on other conditions imposed by the cellular environment.

Keywords

Calmodulin; human Multiprotein Bridging Factor 1; CaM-hMBF1 interaction; IQ-like CaM target; NMR; CaM-agarose affinity chromatography

Abbreviations

CaM: calmodulin; hMBF1: human Multiprotein Bridging Factor 1; hMBF1α_C: human Multiprotein Bridging Factor 1, α isoform, C-terminal domain; hMBF1α_FL human Multiprotein Bridging Factor 1, α isoform, Full Length

1. INTRODUCTION

The intracellular calcium sensor protein Calmodulin (CaM) interacts, through a wide range of binding modes, with a large number of proteins to regulate their biological functions in response to calcium stimuli. The three-dimensional structure of CaM has been well characterized both in the apo and Ca^{2+} forms. The protein consists of two similar globular domains, the N-terminal (1-77) and C-terminal (81-148), each containing two helix-loop-helix (EF-hand) motifs which allow CaM to bind up to four calcium ions per molecule. The domains are connected by a *central linker* which is very flexible in solution, so that the relative orientation of the two domains can change quite readily [1-5]. Within each of the two domains of CaM, significant conformational changes are induced in response to calcium binding. In the absence of calcium each EF-hand motif adopts a "closed conformation", with the helices in an almost antiparallel arrangement and most of the hydrophobic residues shielded from the solvent. Binding of Ca^{2+} normally causes a rearrangement of the helices to the "open conformation", with many hydrophobic residues exposed on the surface of the protein and a large hydrophobic surface created on each domain [6,7]. This conformational switch allows CaM to bind to target proteins through these hydrophobic surfaces, characterized by a high proportion of Met residues. The latter seem to be essential for the unique promiscuous binding behaviour of the protein [8]. On the other hand, the hydrophilic nature of apo-CaM allows the protein to interact with other target proteins in a Ca^{2+}-independent manner.

Although the CaM binding domains of CaM targets share very little sequence identity, many of them show a number of common characteristics. In particular, the highly conserved positions of bulky hydrophobic residues in CaM binding regions has allowed the classification of CaM binding domains into recognition motifs, named **1-14**, **1-10**, **1-16** (with key bulky residues separated by 12, 8 and 14 residues, respectively) and **IQ** [9] (for a continuous update of CaM binding targets, see Calmodulin Target Database, Ontario Cancer Institute: http://calcium.uhnres.utoronto.ca). Atypical sequences that do not conform to any of the above motifs are referred to as **Others**. The IQ class includes two subclasses defined by the presence of the motifs IQ (or *complete IQ*) and IQ-like (or *incomplete IQ*). The first one has the consensus sequence: (F,I,L,V)QXXX(R,K)GXXX(R,K)XX(F,I,L,V,W,Y). However, the G after the first basic

amino acid (almost always R) is not always conserved, nor is the hydrophobic residue in the third position after the second basic amino acid, so the sequence (F,I,L,V)QXXX(R,K)XXXX(R,K) represents a more generalized IQ motif [10]. The IQ motif was first identified as a calcium-independent CaM binding motif [11] but experimental evidence of many exceptions exists [10], now confirmed by the structures of different Ca^{2+}-CaM-IQ complexes [12-15]. The consensus sequence of the IQ-like motifs, (F,I,L,V)QXXX(R,K)XXXXXXXX, differs from the complete IQ for the absence of the second basic residue. The significance of this motif in its CaM binding properties, in some of these proteins, remains to be determined and so far no structural information of CaM/IQ-like complexes is available.

The Multiprotein Bridging Factor 1 (MBF1) is an IQ-like containing protein, highly conserved in eukaryotes and archaea, while none of the sequenced bacterial genomes contains an orthologue of its coding gene [16]. This phylogenetic distribution fits with the functional role of the protein as a transcriptional co-activator [17-26] as eukaryotes and archaea have a high similarity in their transcription initiation complexes, whereas the bacterial systems look very different [16]. The CaM binding motif of this protein is slightly different from the IQ-like sequence, as only two amino acids, instead of three, are interposed between the IQ and the first basic residue (see Table 1). Despite this difference, the sequence seems to be responsible of the interaction with CaM, as revealed by in vivo and in vitro experiments with human and the homologous bovine and rat MBF1 [22,24,27]. Structural data available on hMBF1 and the homologous from Bombyx mori and Tricoderma reesei (PDB ID: 1X57, 2JVL) [28-31] show that the protein contains an N-terminal flexible region and a C-terminal structurally well defined domain with a helix-turn-helix (HTH) motif. The IQ-like sequence belongs to the latter domain and experimental evidence exists of a preferential role of this domain in CaM binding [32]. As the full length protein (here-after called hMBF1α_FL) is not stable in solution, we focused our study on the structured C-terminal domain of hMBF1, α isoform (hMBF1α_C), residues 71-148.

We analysed hMBF1α_C binding to CaM by both NMR and CaM-agarose affinity chromatography. While the latter method gives contradicting results, depending on buffer

conditions, NMR clearly confirms that hMBF1α_C does not show binding capability to CaM, in vitro, both in the absence and in the presence of calcium.

2. MATERIALS AND METHODS

Salts and other chemicals were purchased from Sigma, except where indicated.

hMBF1α_FL cloning and *purification* - The gene coding for hMBF1_FL was amplified from cDNA using two sets of primers with sequences: OMBF1F1 (GGACGGACGCTCGTCTTC) and OMBF1R1 (GCACTGATTTCGAGGCTTTG), for the external primers, and OMBF1F2 (CACCATGGCCGAGAGC) and OMBF1R2 (TCATTTCGCCCTAGGCCCCTT) for the specific ones. The DNA amplified by PCR was purified from agarose gel and cloned in the Gateway plasmid pENTR/TEV/D-TOPO (Invitrogen). The gene was then moved into the Gateway destination vector pETG20A (European Molecular Biology Laboratory, Heidelberg, Germany) to get the plasmid pETG20A_MBF1_FL which encodes for a recombinant TrxHis-tagged hMBF1α_FL. The protein was purified from *E. coli* BL21 culture, induced with 1 mM isopropyl β-thiogalactopiranoside (IPTG; Fermentas). Cells were collected by centrifugation, resuspended in buffer A [50 mM tris-HCl pH 8.0, 50 mM KCl, 1 mM $MgCl_2$, 1 mM Rnase I, 1mM Dnase I, 0.5 mM PMSF, 1 mM Pefabloc, 50 μl/ml of protease inhibitor cocktail] and lysed through a French press operating at 1000 psi. The supernatant obtained by centrifugation was loaded onto a nichel-chelate Hi-Trap column (GE Healthcare) in buffer B [50 mM Tris-HCl pH 8.0, 50 mM KCl]. Elution of the protein was carried out with buffer C [50 mM Tris-HCl pH 8.0, 50 mM KCl, 50 mM EDTA]. The protein was further purified by cation-exchange chromatography (Macro-Prep High S Support, Bio-Rad) in buffer D [20 mM Tris pH 8.8, 50 mM KCl] and eluted with a linear KCl gradient. The His-tag was cleaved with ActTev protease (Invitrogen) and two cation-exchange chromatographies were performed after cleavage, at pH 5.5 with buffer F [20 mM Mes, 50 mM KCl], and at pH 8.8 with buffer G [20 mM Tris pH 8.8, 50 mM KCl], using for the latter a BIO PolyMa SCX column (Supelco). Protein elution was obtained with linear KCl gradient. Samples for NMR experiments were exchanged against buffer H [20 mM Hepes, 200 mM KCl, pH 7.4] and their concentration was

determined spectrophotometrically using as extinction coefficient the value ε_{280} = 12660 $M^{-1}cm^{-1}$, calculated by the ProtParam tool [33] (http://www.expasy.org/tools/protparam.html). Samples purity was checked by SDS-page.

hMBF1α_C cloning and *purification* - The gene coding for amino acids 71-148 of hMBF1 was amplified from plasmid pETG20A_MBF1_FL using primers oMBF1F (5'GACAGGGTGACCCTGGAGGT3') and oMBF1R (5'TCATTTCGCCCTAGGCCC3'), and cloned in plasmid pQE30Xa (QIAGEN) within the StuI site. Recombinant His-tagged hMBF1α_C was purified from *E. coli* XL1 Blue culture after 4 hrs from induction with 1 mM IPTG. The purification protocol was the same as for the full length protein, but Factor Xa (Amersham Biosciences) was used to cleave the His-tag. Protein concentration was determined spectrophotometrically using as extinction coefficient the value ε_{280} = 1490 $M^{-1}cm^{-1}$, calculated by the ProtParam tool [33]. Protein molecular mass was confirmed by MALDI-TOF spectrometry performed by the CIRB-CRBA Proteomic Facility Services -LaP- c/o CIRB-CRBA University Hospital St Orsola Malpighi, Bologna. The value obtained, 8558 Da is comparable, within experimental error, to that calculated by ProtParam tool [33], 8560 Da, for the protein sequence:
DRVTLEVGKVIQQGRQSKGLTQKDLATKINEKPQVIADYESGRAIPNNQVLGKIE
RAIGLKLRGKDIGKPIEKGPRAK.

Purification of ^{15}N labelled hMBF1α_FL and ^{15}N labelled hMBF1α_C - ^{15}N labelled hMBF1α_FL and hMBF1α_C were obtained from bacterial growth in M9 minimal medium containing $(^{15}NH_4)_2SO_4$ as the only nitrogen source. The expression was increased using a 1:1 (v/v) LB pre-inoculum, adding 1mM IPTG after 2 hours from inoculum and collecting the cells after an over night growth. The purification protocol was the same used for the unlabelled proteins.

CaM preparation – Unlabelled and ^{15}N labelled N60D CaM were purchased from ProtEra srl (Florence, Italy, http://www.proterasrl.com/). Calcium N60D CaM sample was prepared by NMR titration of apo CaM with calcium until the protein was completely in the calcium-bound form [34,35]. The buffer used for the NMR experiments was buffer H [20 mM Hepes, 200 mM KCl, pH 7.4].

NMR experiments - All the NMR titration experiments were performed at Bruker 700 MHz spectrometer equipped with a cryo-probe at 298 K. CaM concentration was around 0.65 mM. ^{1}H-^{15}N HSQC spectra were performed after every addition of 0.5 equivalent hMBF1α_C until the 1:3 ratio of CaM (apo and Ca^{2+} bound) to hMBF1α_C was reached. ^{15}N labelled hMBF1α_C (around 0.4 mM) was also used to acquire the ^{1}H-^{15}N HSQC spectra without and with calcium(II) addition until 4 equivalents, to exclude the possibility of interaction between hMBF1α_C and calcium. Unlabelled apo and holo forms of CaM were titrated into ^{15}N labelled hMBF1α_C in steps of 0.5 equivalents until the 1:1.5 ratio of hMBF1α_C to CaM was reached.

hMBF1α_C binding to CaM-agarose - Binding of hMBF1α_C to a CaM-agarose (Sigma) column was checked in the following sample/column equilibration buffer conditions: A) 50 mM Tris, 5 mM $CaCl_2$, pH 7.4; B) 50 mM Tris, 5 mM EGTA, pH 7.4; C) 20 mM Hepes, 5 mM $CaCl_2$, pH 7.4; D) 20 mM Hepes, 5 mM EGTA, pH 7.4; E) and F) 20 mM Hepes, pH 7.4 metal free; G) 20 mM Hepes, 200 mM KCl, pH 7.4. For any experiment, 250 μg of pure protein were loaded on a 5 ml CaM-agarose column. After a long washing step with the same column equilibration buffer, elution was performed, respectively, with buffers: A') 50 mM Tris, 10 mM EGTA, pH 7.4; B') 50 mM Tris, 10 mM $CaCl_2$, pH 7.4; C') 20 mM Hepes, 10 mM EGTA, pH 7.4; D') 20 mM Hepes, 200 mM KCl, 10 mM $CaCl_2$, pH 7.4; E') 20 mM Hepes, 10 mM $CaCl_2$, pH 7.4 metal free; F') 20 mM Hepes, 200 mM KCl, metal free; G') 20 mM Hepes, 200 mM KCl, 10 mM $CaCl_2$, pH 7.4. The loading and washing fractions were pooled and then concentrated to 250 μl, as well as the elution fractions. An aliquot of both samples (the loading/washing sample and the elution sample) was loaded on SDS-PAGE to check the presence of the protein after passing through the column, in all the experimental conditions tested.

3. RESULTS

Preliminary NMR experiments on hMBF1α – We first produced hMBF1 in its full length sequence, as a fusion protein with TrxHis tag. After excision of the tag the protein looked stable in solution (Figure 1A) but it progressively underwent N-terminal degradation, in a time period of a few days, variable as a function of temperature (being

longer at low temperatures), and of the presence of protease inhibitors. This behaviour suggests that the N-terminal degradation might depend on the action of bacterial proteases and might be related to the lack of a tertiary structure of the N-terminal region, as shown by the ^{1}H-^{15}N HSQC spectrum of ^{15}N hMBF1α_FL (Figure 1B). In that spectrum, approximately 70 resolved peaks were observed (instead of 148), in addition to severely overlapping signals in the spectral region 8.0-8.5 p.p.m. for ^{1}H and 119-124 p.p.m. for ^{15}N, in which backbone signals of amide groups in random coil conformation are usually observed. The C-terminal domain alone, with a sequence corresponding to amino acids 71-148 of hMBF1α (Figure 1A), gave a pattern of peaks essentially identical to the well dispersed peaks in the spectrum of the full-length protein (Figure 1B). The assigned backbone chemical shifts of $^{15}N^{13}C$ labelled hMBF1α_C are reported in Table 2 (supplementary material). These results suggest that hMBF1α has both flexible (in the N-terminal domain) and well-structured (in the C-terminal domain) parts in solution. These data are in agreement with those reported for Bombyx mori MBF1 [28], and explain why the only high resolution structures available till now are referred to the MBF1 C-terminal domain, as in the case of the human (PDB ID: 1X57, unpublished) [31] and Tricoderma reesei orthologs (PDB ID:2JVL) [30,31].

Monitoring CaM-hMBF1α_C binding by NMR - The binding of hMBF1α_C to CaM was monitored through the ^{1}H-^{15}N HSQC spectrum of ^{15}N-labelled CaM in the presence of increasing amounts of hMBF1α_C up to 1:3 equivalents (Figure 2). No change in the spectrum occurred in the absence of calcium(II) in solution, indicating that hMBF1α_C does not bind apo-CaM (Figure 2A). Analogously, the ^{1}H-^{15}N HSQC spectrum of Ca^{2+}-CaM did not change appreciably by adding hMBF1α_C, indicating again that no relevant interaction between the two proteins occurs (Figure 2B). The same result was obtained by monitoring the ^{1}H-^{15}N HSQC spectrum of ^{15}N-labelled hMBF1α_C in the presence of increasing amounts of apoCaM or holoCaM (Figure 3). The ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled hMBF1α_C in fact did not change either after the addition of calcium ions in solution, of apo CaM (Figure 3A) or of calcium-bound CaM (Figure 3B).

hMBF1α_C binding to CaM-agarose - To check the interaction of CaM with hMBF1α_C we used a CaM-agarose affinity chromatography column (Sigma), first

under the same experimental conditions used for hMBF1α [22,24], for the homologous bovine MBF1 [27], and for a 20 aa peptide spanning the IQ motif of rat MBF1 [27]. The recombinant hMBF1α_C was loaded on the column in the presence of calcium and in the presence of EGTA. The elution step was performed with a buffer containing EGTA for the first experiment and calcium for the second one. Figure 4 shows that hMBF1α_C does not bind to the column in the presence of calcium both with Tris or Hepes buffer (lanes A-A' and C-C' of Figure 4) but binds in the presence of EGTA (i.e. in the absence of free calcium), independently from the buffer (lanes B-B' and D-D' of Figure 4), confirming what previously reported in the literature [22,24]. The same result was obtained using a metal free buffer (i.e. without both calcium and EGTA; lane E-E' of Figure 4) apparently confirming that hMBF1α_C binds apo-CaM and not its calcium bound form. However, once bound to the column, the protein can be eluted not only with $CaCl_2$ but also with KCl (the latter in the absence of calcium, see lane F-F' of Figure 4). Besides, the protein is not retained by the column when loaded in the presence of KCl (lane G-G' of Figure 4). These results suggest that binding of hMBF1α_C to the column is not dependent on the calcium-free state of CaM but depends on ionic interactions with the activated matrix, which can be prevented by the presence of salts.

4. DISCUSSION

Information available in literature about the interaction of CaM with MBF1 are referred mainly to the application of CaM-agarose Affinity Chromatography, using a column where CaM is immobilized by cyanogen bromide activation to a 4% beaded agarose matrix. Results obtained by this technique, on the recombinant full length hMBF1α [22,24] and the homologous full length bovine MBF1 [27], indicate that the two proteins interact in the apo form. Analogous result was obtained for a 20 aa peptide spanning the IQ motif of rat MBF1 [27]. On the contrary, another group reports a calcium-mediated interaction between CaM and hMBF1 determined using an in vitro binding assay with ^{35}S-GST fusion CaM and GST fusion hMBF1 [32]. In the latter work a stronger binding of GST-hMBF1 to GST-CaM was observed in the presence of Ca^{2+} than without, and an N-terminally truncated hMBF1, missing amino acids 1 to 77,

produced as GST-fusion protein, bound more strongly CaM than wild-type GST-hMBF1 [32].

In order to make progress in the understanding of the CaM-hMBF1 interaction we monitored by NMR the formation of the complex between the two proteins, and compared the results with those obtained with CaM-agarose affinity chromatography in different experimental conditions. In order to exclude a possible influence of not native amino acids, proteins were purified without any peptide or protein tag. Experiments were done with both the full length form of hMBF1 (data not shown), which contains an N-terminal unstructured domain, and the C-terminal domain alone (hMBF1α_C), which is structured and stable in solution. The C-terminal domain of hMBF1α contains the IQ-like motif supposed to be responsible for CaM binding, and evidences exist in the literature for a preferential role of this domain in CaM complex formation [32]. Both form of the protein (FL and C-terminal domain) gave equivalent results by NMR and CaM-agarose affinity chromatography.

In particular, the ^{1}H-^{15}N HSQC spectrum of both ^{15}N-labelled CaM in the presence of increasing amounts of hMBF1α_C and ^{15}N-labelled hMBF1α_C in the presence of increasing amounts of CaM clearly showed that no sizable interaction between the two proteins occurs, both in the presence and in the absence of calcium, in a buffer containing KCl at physiological concentration. On the other hand, we verified that the presence of KCl (in a calcium free buffer), prevents binding of hMBF1α_C to the CaM-agarose column, and that when bound to the column the protein can be eluted not only with addition of calcium but also with KCl (in the absence of calcium). On the contrary, without salts (in a buffer with or without EGTA), hMBF1α_C binds perfectly to the column as reported in previous studies [22,24,27]. Altogether these results suggest that binding of hMBF1α_C to the column is not dependent on the calcium-free state of CaM but rather depends on nonspecific interactions between hMBF1α_C and the CaM or the activated matrix of the column, which can be prevented by the presence of salt, be it KCl or $CaCl_2$. Right now it is not possible to know if CaM or the matrix generate nonspecific binding sites for hMBF1α_C, when salts are omitted from the solution. In this respect, the control experiment with the matrix alone, made by the other groups who have used the CaM-agarose affinity chromatography to study the interaction between MBF1 and

CaM [22,24,27], and which gave no interaction of MBF1 with the matrix, is not sufficient to clarify this point. In fact, we cannot exclude that nonspecific binding sites in the agarose matrix are generated during the activation of the resin for CaM-binding.

These results are particularly meaningful because they point out that a single experimental approach may not be sufficient to characterize protein-protein interactions and that experimental conditions far from physiological ones (as absence of potassium chloride in sample solution) are critical points that can generate false results. The nonspecific binding of MBF1α to the CaM-column, observed in the present and previous studies, suggest that, even though affinity chromatography with immobilized proteins is considered the method of choice both for analytical purposes and for quantitative assessment of such interactions, it should be regarded as a preliminary step whose results need to be confirmed, when possible, by spectroscopic experiments able to directly detect perturbations due to the presence of the interacting protein. This caveat possibly holds also for peptides that bind to CaM linked to beaded agarose matrix, for which innumerable works referenced in the literature make the assumption that most CaM binding proteins only interact with the immobilized CaM and not with the column matrix. For this reason, particular attention must be paid when CaM, as well as other immobilized proteins are used to measure their affinity with putative partners.

Results so far obtained by in vitro experiments, do not exclude that an interaction between the two proteins exists in vivo, as suggested by Ballabio and co-workers [24] for hMBF1 of endothelial cells (named EDF1), on the base of immunoprecipitation assay on HUVEC cells. The authors speculate that EDF1 can serve, in the cytoplasm of endothelial cells, as a regulator of CaM availability and, consequently, of the activation of enzymes necessary for the maintenance of the vascular integrity. In the light of our in vitro experiments it is likely that if an interaction between the two proteins exists in vivo, it involves a modified form of CaM, or of hMBF1 or of both proteins. This would not be surprising as native protein assemblages are formed in a complex environment, and can be regulated by different mechanisms such as folding, modification, limited proteolysis, transportation to specific cellular compartments, and assembly with non-proteinous co-factors. In conclusion, more experimental data are needed to elucidate the molecular

mechanism of MBF1/CaM interaction and its putative role in the functional network created inside the cells by CaM binding to its many competing targets.

5. AKNOWLEDGMENTS

*Jing Yuan is acknowledged for preliminary NMR experiments. This work has been supported by Ente Cassa di Risparmio di Firenze, Consorzio Interuniversitario Risonanze Magnetiche di Metalloproteine Paramagnetiche (CIRMMP), Polo Scientifico-Didattico di Cesena, MIUR-FIRB contracts RBLA032ZM7, RBIP06LSS2, RBRN07BMCT, PRIN 2007M5MWM9 and by the EC contracts EU-NMR 026145, SPINE2-COMPLEXES 031220, and LSHB-CT-2005-019102.

6. REFERENCE LIST

[1] G. Barbato, M. Ikura, L.E. Kay, R.W. Pastor, A. Bax, Biochemistry 31 (1992) 5269-5278.

[2] F. Capozzi, F. Casadei, C. Luchinat, J. Biol. Inorg. Chem. 11 (2006) 949-962.

[3] N. Tjandra, H. Kuboniwa, H. Ren, A. Bax, Eur. J. Biochem. 230 (1995) 1014-1024.

[4] D. van der Spoel, B.L. de Groot, S. Hayward, H.J. Berendsen, H.J. Vogel, Protein Sci. 5 (1996) 2044-2053.

[5] V. Borsi, C. Luchinat, G. Parigi, Biophys. J. 97 (2009) 1765-1771.

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

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

[8] A.P. Yamniuk, H.J. Vogel, Mol. Biotechnol. 27 (2004) 33-57.

[9] K.L.Yap, J. Kim, K. Truong, M. Sherman, T. Yuan, M. Ikura, J. Struct. Funct. Genomics. 1 (2000) 8-14.

[10] M. Bahler, A. Rhoads, FEBS Lett. 513 (2002) 107-113.

[11] A.R. Rhoads, F. Friedberg, FASEB J. 11 (1997) 331-340.

[12] J.L. Fallon, D.B. Halling, S.L. Hamilton, F.A. Quiocho, Structure 13 (2005) 1881-1886.

[13] E.Y. Kim, C.H. Rumpf, Y. Fujiwara, E.S. Cooley, F. Van Petegem, D.L. Minor, Structure 16 (2008) 1455-1467.

[14] F. Van Petegem, F.C. Chatelain, D.L. Minor Jr, Nat. Struct. Mol. Biol. 12 (2005) 1108-1115.

[15] M.X. Mori, C.W. Vander Kooi, D.J. Leahy, D.T. Yue, Structure 16 (2008) 607-620.

[16] B. de Koning, F. Blombach, H. Wu, S.J. Brouns, J. Van der Oost, Biochem. Soc. Trans. 37 (2009) 52-57.

[17] F.Q. Li, H. Ueda, S. Hirose, Mol. Cell. Biol. 14 (1994) 3013-3021.

[18] F.Q. Li, K. Takemaru, M. Goto, H. Ueda, H. Handa, S. Hirose, Genes Cells. 2 (1997) 143-153.

[19] K. Takemaru, F.Q. Li, H. Ueda, S. Hirose, Proc. Natl. Acad. Sci. USA 94 (1997) 7251-7256.

[20] K. Takemaru, S. Harashima, H. Ueda, S. Hirose, Mol. Cell. Biol. 18 (1998) 4971-4976.

[21] Y. Kabe, M. Goto, D. Shima, T. Imai, T. Wada, K. Morohashi, M. Shirakawa, S. Hirose, H. Handa, J. Biol. Chem. 274 (1999) 34196-34202.

[22] M. Mariotti, L. De Benedictis, E. Avon, J.A. Maier, J. Biol. Chem. 275 (2000) 24047-24051.

[23] C. Brendel, L. Gelman, J. Auwerx, Mol. Endocrinol. 16 (2002) 1367-1377.

[24] E. Ballabio, M. Mariotti, L. De Benedictis, J.A. Maier, Cell. Mol. Life Sci. 61 (2004) 1069-1074.

[25] J.J. Millership, P. Waghela, X. Cai, A. Cockerham, G. Zhu, Microbiology. 150 (2004) 1207-1213.

[26] D.P. Arce, C. Tonon, M.E. Zanetti, A.V. Godoy, S. Hirose, C.A. Casalongue, J. Biochem. Mol. Biol. 39 (2006) 355-360.

[27] M.L. Smith, R.A. Johanson, K.E. Rogers, P.D. Coleman, J.R. Slemmon, Mol. Brain Res. 62 (1998) 12-24.

[28] M. Mishima, J. Ozaki, T. Ikegami, Y. Kabe, M. Goto, H. Ueda, S. Hirose, H. Handa, M. Shirakawa, J. Biomol. NMR 14 (1999) 373-376.

[29] J. Ozaki, K.I. Takemaru, T. Ikegami, M. Mishima, H. Ueda, S. Hirose, Y. Kabe, H. Handa, M. Shirakawa, Genes Cells 4 (1999) 415-424.

[30] R.K. Salinas, C.M. Camilo, S. Tomaselli, E.Y. Valencia, C.S. Farah, H. El Dorry, F.S. Chambergo, Proteins 75 (2009) 518-523.

[31] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucleic Acids Res. 28 (2000) 235-242.

[32] P.K. Busk, L. Wulf-Andersen, C.C. Strom, M. Enevoldsen, K. Thirstrup, S. Haunso, S.P. Sheikh, Exp. Cell Res. 286 (2003) 102-114.

[33] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein Identification and Analysis Tools on the ExPASy Server, in: J.M. Walker (ed.), The Proteomics Protocols Handbook, Humana Press, 2005, pp. 571-607.

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

[35] I. Bertini, I. Gelis, N. Katsaros, C. Luchinat, A. Provenzali, Biochemistry 42 (2003) 8011-8021.

FIGURE CAPTIONS

Figure 1: A) SDS-page of recombinant hMBF1α_FL and hMBF1α_C; B) overlapped ^{1}H-^{15}N HSQC spectra of hMBF1α_FL (red) and hMBF1α_C (green).

Figure 2: ^{1}H-^{15}N HSQC of (A) apo-^{15}NCaM (red) and apo-^{15}NCaM:hMBF1α_C = 1:3 (green); and (B) Ca_4-^{15}NCaM (red) and Ca_4-^{15}NCaM:hMBF1α_C = 1:3 (green).

Figure 3: ^{1}H-^{15}N HSQC of (A) ^{15}NMBF1 before and after adding CaM and (B) ^{15}NMBF1 before and after adding Ca_4-CaM.

Figure 4: SDS-page of samples obtained by CaM-agarose affinity chromatography. Pure hMBF1α_C was loaded on CaM-agarose column, washed with loading/washing buffers (LW) A to G and eluted with elution buffers (E) A' to G'. Composition of buffers is specified on the side of the SDS-page. An aliquot of pure hMBF1α_C was loaded as control (CT).

Figure 1:

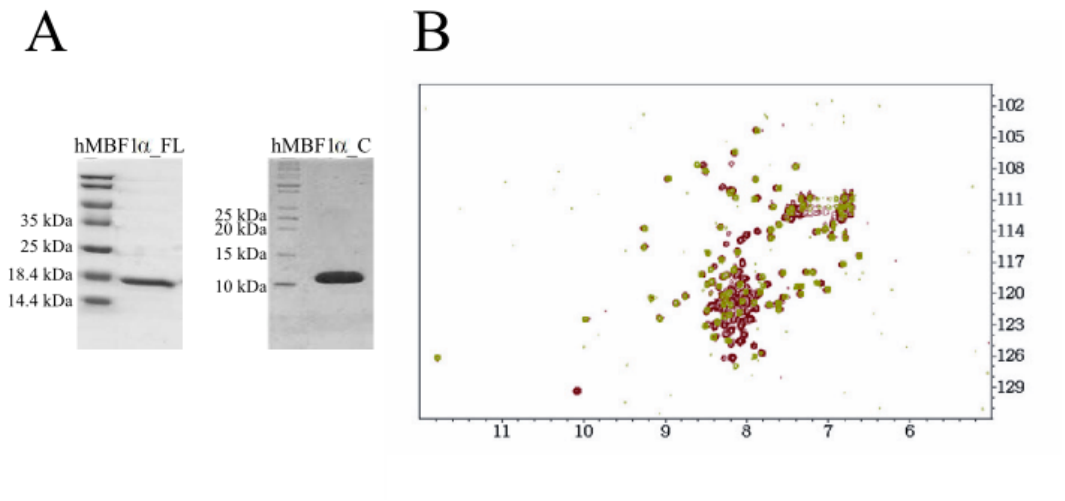


Figure 2:

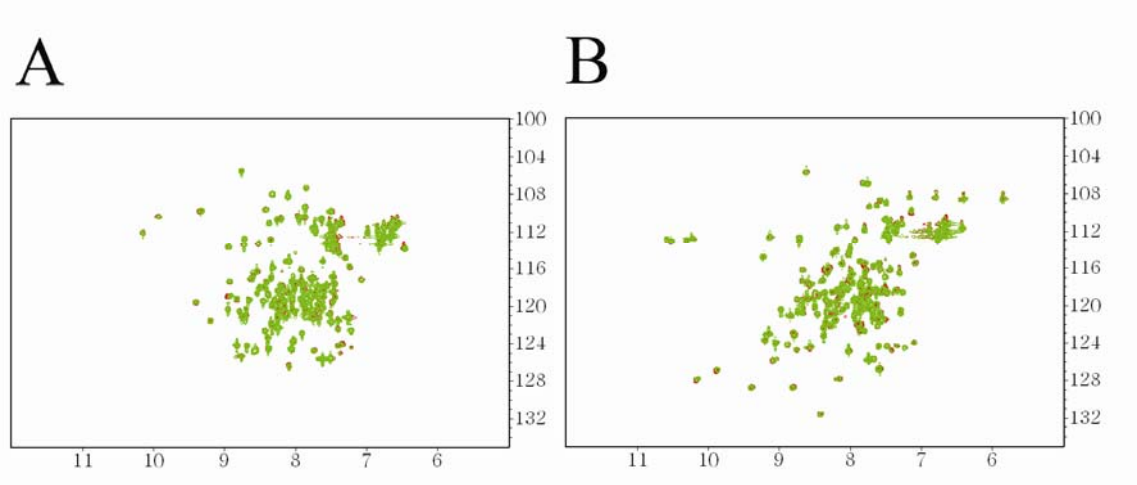


Figure 3:

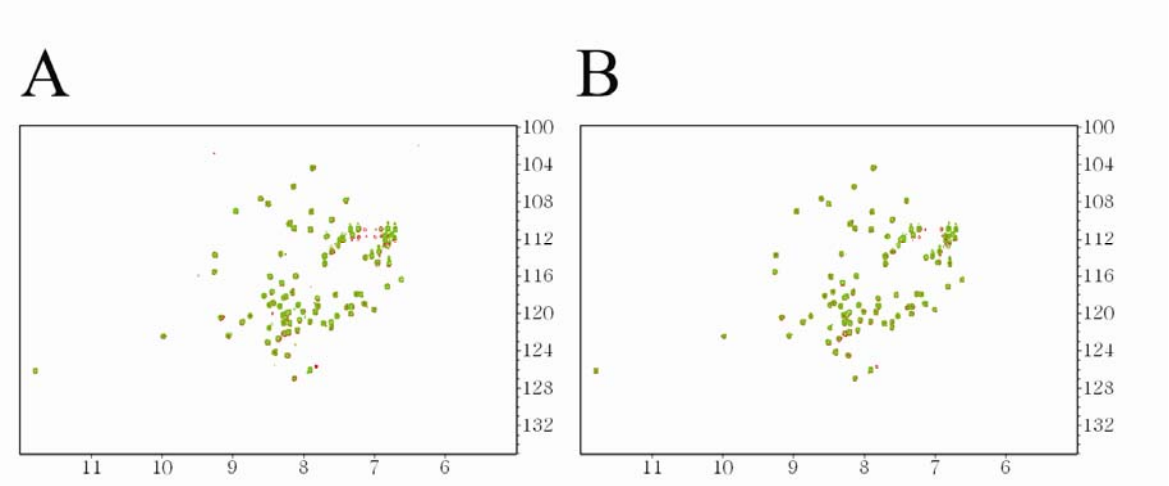
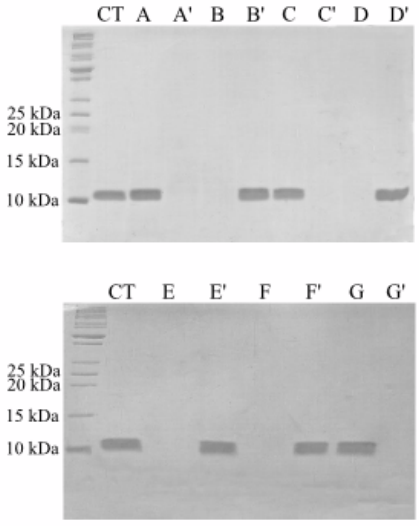


Figure 4:



Lane	Buffer applied to the CaM-agarose column
A	LW 50 mM Tris, 5 mM $CaCl_2$, pH 7,4
A'	E 50 mM Tris, 10 mM EGTA, pH 7.4
B	LW 50 mM Tris, 5 mM EGTA, pH 7.4
B'	E 50 mM Tris, 10 mM $CaCl_2$, pH 7.4
C	LW 20 mM Hepes, 5 mM $CaCl_2$, pH 7.4
C'	E 20 mM Hepes, 10 mM EGTA, pH 7.4
D	LW 20 mM Hepes, 5 mM EGTA, pH 7.4
D'	E 20 mM Hepes, 200 mM KCl, 10 mM $CaCl_2$, pH 7.4
E	LW 20 mM Hepes, pH 7.4 metal free
E'	E 20 mM Hepes, 10 mM $CaCl_2$, pH 7.4 metal free
F	LW 20 mM Hepes, pH 7.4 metal free
F'	E 20 mM Hepes, 200 mM KCl, pH 7.4 metal free
G	LW 20 mM Hepes, 200 mM KCl pH 7.4
G'	E 20 mM Hepes, 10 mM $CaCl_2$, 200 mM KCl, pH 7.4

Table 1: Amino acid sequence of IQ and IQ-like motifs

Motif	Sequence
IQ	**I basic residue** ↓ **II basic residue** ↓ (F,I,L,V)QXXX(R,K)GXXX(R,K)XX(F,I,L,V,W,Y)
generalized IQ	(F,I,L,V)QXXX(R,K)XXXX(R,K)XXX
IQ-like	(F,I,L,V)QXXX(R,K)XXXXXXXX
hMBF1α IQ-sequence	IQQGRQSKGLTQK

Table 2 The assigned backbone chemical shifts of $^{15}N^{13}C$ labelled hMBF1α_C

Res. Num.	Res. type	N	C	C_α	C_β	H_N	H_α	H_β/Q_β
76	ASP	122.953	172.883	51.462	37.925	7.938	4.224	
77	ARG	119.666	173.83	52.697	28.18	7.864	4.102	1.433
78	VAL	121.991		59.761	29.467	8.052	3.726	1.634
79	THR							
80	LEU							
81	GLU							
82	VAL		174.533					
83	GLY	107.04	172.836	45.804		8.284	3.853 3.468	
84	LYS	119.395	176.317	55.614	28.947	7.321	3.889	1.665
85	VAL	121.449	175.707	63.425	27.998	7.446	3.356	1.842
86	ILE	119.697	174.098	63.369	34.759	8.035	3.144	1.719
87	GLN	118.186	175.241	57.007	25.982	7.688		
88	GLN	118.186	176.961	55.838	25.219	8.136	3.734	1.86
89	GLY	108.826	173.428	44.394		8.81	3.307 3.646	
90	ARG	122.309	175.986	58.029	26.0	8.92	3.582	1.525
91	GLN	119.104	177.812	56.313	24.721	8.18	3.867	1.986 1.839
92	SER	116.701	172.722	58.657	59.853	8.165	4.02	3.794
93	LYS	117.663	173.592	51.87	30.354	7.01	4.174	1.783 1.555
94	GLY	109.099	171.36	43.3		7.718	3.608	
95	LEU	117.658	175.546	50.21	41.039	7.388	4.635	1.131 1.434
96	THR	115.626	173.202	58.453	68.19	9.126	4.081	4.494
97	GLN	120.179	175.249	58.44	24.912	9.052	3.34	1.902
98	LYS	117.977	175.263	56.933	28.976	8.129	3.538	1.519
99	ASP	120.14	176.767	54.209	38.827	7.923	4.048	2.345
100	LEU	119.966	173.594	54.808	38.204	7.864	3.655	1.654 1.443
101	ALA	121.275	176.302	52.859	14.453	8.049	3.293	1.149
102	THR	113.77	174.604	63.378	66.012	8.202	3.69	3.988
103	LYS	120.96	175.286	56.271	29.54	7.549	3.785	1.687
104	ILE	107.595	171.36	57.364	35.72	7.241	4.372	1.809
105	ASN	119.768	170.919	51.285	33.835	7.652	4.058	2.857 2.4
106	GLU	115.352	172.377	50.506	30.192	7.923	4.371	1.848 0.824
107	LYS	120.628		52.013	28.518	8.744	4.176	1.354
108	PRO		175.207					
109	GLN	116.056	174.102	55.225	25.174	8.29	3.524	1.835 1.529
110	VAL	117.639	174.922	62.906	28.492	7.16	3.382	1.859
111	ILE	115.899	174.92	58.818	31.779	6.398	3.229	1.835
112	ALA	120.55	178.254	52.362	14.371	7.435	3.695	1.119
113	ASP	119.847	176.072	54.212	36.341	8.26	4.043	2.627 2.038

114	TYR	121.937	176.512	57.25		8.378		
115	GLU	116.857	174.843	56.5	26.331	8.106	3.873	1.647
116	SER	110.721	172.722	55.371	61.462	7.63	4.173	3.673
117	GLY	110.135	171.165	42.883		7.527	3.733	
118	ARG	118.088	173.038	54.271	28.91	7.82	3.923	1.454 1.252
119	ALA	118.967	172.833	47.955	17.731	6.801	3.987	0.582
120	ILE	123.138		54.985	35.289	8.354	4.118	1.5
121	PRO		172.645					
122	ASN	119.006		48.957	37.458	7.637		
123	ASN		174.533	53.939	35.37			
124	GLN	120.66	176.495	56.427	24.89	8.081	3.843	1.929
125	VAL	121.937	175.472	64.193	28.387	7.952	3.131	1.845
126	LEU	118.547	176.182	55.536	38.915	8.326	3.485	1.533 0.954
127	GLY	104.234	173.197	44.014		7.725	3.582	
128	LYS	119.768	177.275	57.335	29.763	7.168	3.794	1.413
129	ILE	119.847	174.461	63.898	34.453	8.07	3.094	1.517
130	GLU	120.003	176.218	56.544	27.547	8.598	3.799	2.085 1.51
131	ARG	113.789	175.244	55.716	26.926	7.021	3.61	1.574
132	ALA	120.159	176.412	51.298	16.247	7.358	3.925	1.012
133	ILE	106.188	172.413	58.26	35.877	7.989	4.139	1.594
134	GLY	110.682	171.627	43.56		7.982	3.752 3.455	
135	LEU	119.534	171.234	49.893	43.827	6.823	4.607	0.944 0.656
136	LYS	117.619	174.998	54.158	30.179	8.392	3.573	1.391
137	LEU	121.820	171.045	50.902	41.269	9.786	4.407	1.575 1.207
138	ARG	114.336	172.854	51.345	30.915	6.764	4.227	1.527 1.41
139	GLY	108.025	171.832	41.893		8.429	3.492 3.85	
140	LYS	119.886	174.4	54.849	29.269	8.37	3.85	1.496
141	ASP	118.772	170.844	50.759	37.376	8.356	4.435	2.59 2.275
142	ILE	116.974	173.822	59.865	34.781	6.647	3.056	1.307
143	GLY	113.867	170.34	41.976		9.126	4.022 3.108	
144	LYS	120.687		51.374	28.46	7.762	4.317	1.586
145	PRO		174.252					
146	ILE	118.791	172.962	58.642	35.7	7.109	3.603	1.27
147	GLU	124.214	173.309	53.242	27.11	8.31	3.958	1.657
148	LYS	122.914	173.874	53.421	30.057	8.334	3.986	1.489
149	GLY	110.155		41.419		8.136	3.77	
150	PRO		174.22					
151	ARG	121.165	173.123	52.874	27.561	8.28	4.004	1.51
152	ALA	126.314	173.891	49.489	16.085	8.114	3.959	1.082
153	LYS	125.943		54.824	30.385	7.791	3.784	1.476 1.388

3.4

Improving the maximum occurrence analysis of calmodulin conformations in solution by placing paramagnetic ions in both protein domains

Ivano Bertini, [1, 2*] Soumyasri Das Gupta, [1] Xiaoyu Hu, [1] Peter H. J. Keizers, [3] Claudio Luchinat, [1, 2] Malini Nagulapalli, [1]Giacomo Parigi, [1, 2] Marcellus Ubbink[3]

[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]Leiden institute of chemistry, Gorlaeus laboratories, Leiden University, Post office box 9502, 2300 RA Leiden, Netherlands.

Improving the maximum occurrence analysis of calmodulin conformations in solution by placing paramagnetic ions in both protein domains

Introduction

Conformational flexibility is often a crucial feature for proteins to perform their function in solution.[1;2] During biological processes, different kinds of conformational changes may occur such as side chain rotations, loop motions, interdomain reorientations, intermolecular rearrangements, random-coil motions in unfolded proteins or protein regions. NMR spectroscopy has long been used for structural and dynamic studies of proteins in solution. In NMR experiments, solution conditions such as temperature, pH and salt concentration can be adjusted to closely mimic the physiological fluid where the protein performs its function. Most protein dynamic studies are focused on the analysis of relaxation data (R_1, R_2 and NOE) and provide information on the protein tumbling times and on the presence of local motions.[3;4] Paramagnetism-based restraints have been shown to monitor the presence of conformational rearrangements among protein domains,[5] to detect the presence of minor interconverting conformations,[6-8] to determine whether regions in the conformational space must be occupied or cannot be occupied by protein complexes,[9;10] and to provide information on the maximum occurrence (MO) of any conformation that is sterically allowed.[11]

The MO strategy is focused on determining the maximum weight that any given conformation can have in any conformational ensemble in agreement with all available experimental data obtained, e.g., through solution NMR or small angle scattering (SAS) measurements.[11-13] These measurements in fact provide weighted averages over all the conformations experienced by the system. They cannot be used to recover the actual protein conformational ensemble, but do provide the maximum percent of time that a system can spend in any conformation.

Calmodulin (CaM) is a calcium(II) EF-hand protein, which contains two similar

globular domains connected by a flexible linker.[14] This structural feature makes it easy for the two domains to adopt a variety of different orientations with respect to one another. In order to describe the interdomain conformational variability, lanthanide ions were used as paramagnetic probes, and NMR experiments were performed to obtain pseudocontact shifts (pcs) and residual dipolar couplings (rdc).[5;12] These data were used as restraints to calculate the conformations with largest MO and to analyze the different MO of the possible protein conformations. The N60D CaM mutant was actually used to selectively substitute the Ca^{2+} ion located in the second binding site of the N-terminal domain with a paramagnetic lanthanide ion (Tb^{3+}, Tm^{3+} or Dy^{3+}).[15] The same analysis was performed to detect the conformational heterogeneity of CaM bound to α-synuclein[12] or to a peptide from the myelin basic protein (MBP, unpublished results from the CERM laboratory). Pcs and rdc restraints were also used to detect slight conformational changes in CaM when bound to peptides representing the interaction sequence of two protein partners, the death-associated protein kinase (DAPk) and the DAPk-related protein 1 (DRP-1), on passing from crystal to solution.[16] The results suggest that the two domains are relatively flexible with respect to one another in free CaM and that mobility changes after target peptide binding.[17]

The advantage provided by the paramagnetism-based restraints is based on the possibility to retrieve the magnetic susceptibility anisotropy tensors of the different metals from the pcs values collected for nuclei belonging to the same domain where the metal ion is coordinated (the N-terminal domain in this case). The pcs and rdc values collected for the nuclei belonging to the other domain act as reporters of the interdomain conformational variability. In fact, they are the weighted average of the pcs and rdc values corresponding to all sampled conformations, and such values are determined by the same magnetic susceptibility anisotropy tensors calculated for the metal bearing domain which also act as orientation tensors.

The MO value calculated for each conformation decreases towards the actual probability when the number of independent experimental restraints is increased. MO values of less probable conformations are expected to decrease more than those of the most probable conformations. The resulting larger spreading of the MO values calculated for the different conformations likely permits to better identify those with largest probability.

In this work we analyze how MO values are affected by the availability of restraints provided by paramagnetic metal ions located in both domains of CaM. Pcs and rdc collected for the N-terminal domain, when the paramagnetic metal is placed in the C-terminal domain of CaM, actually represent independent information and provide a different perspective on the protein conformational variability. Therefore, when the paramagnetic restraints obtained from both the metal positions are used together, the difference in the MO values of highly occurring and lowly occurring conformations is expected to increase. Furthermore, besides providing additional information on the relative position of the domains, the addition of these restraints could also remove some of the possible "ghost" solutions determined by the mathematical form of the pcs and rdc equations.[12;13]

In order to place a metal ion in the C-terminal domain of CaM, the Caged Lanthanide NMR Probe 5 (CLaNP-5)[18] was attached to the H107C/N111C CaM mutant. This tag was chosen because it can bind rigidly to the protein backbone through two cysteine residues.[19] Rigid binding is essential to obtain the correct magnetic susceptibility anisotropy tensor from the pcs of the attached domain and to easily interpret the pcs of the other domain.[20]

Materials and Methods

Protein preparation

^{15}N labeled N60D CaM was purchased from ProtEra s.r.l. (Florence, Italy, www.proterasrl.com). The NMR samples were prepared in 20 mM MES, 200 mM KCl, pH 6.8. For the attachment of the ClaNP-5 tag, the H107C/N111C mutations were introduced in wild type CaM via site-directed mutagenesis. ^{15}N labeled his-tagged H107C/N111C CaM was expressed in E. coli BL21(DE3) Gold cells and purified with Ni-NTA column and size exclusion chromatography in the same buffer as the N60D mutant. The entire process of tag attachment was performed under reducing conditions. The Ca^{2+}-CaM mutant was incubated with 5 mM DTT for 30 mins to reduce all possible disulfide bridges and ensure that the protein existed in monomeric state. DTT was then washed out under reducing conditions. The protein was diluted to a concentration of 30 μM. Seven equivalents of Ln^{3+}-loaded CLaNP-5 (Ln^{3+} = Lu^{3+}, Yb^{3+} and Tm^{3+}) was added to it. The mixture was incubated

overnight at 4 °C for the reaction to reach completion. To separate the tagged monomeric protein from aggregates and free tag present in solution, a purification was performed using a Superdex 200 gel filtration column. Approximately three-fourth of the total protein was found to be monomeric and reacted while the remaining one-fourth formed aggregates. In all the above steps of tagging, the CaM mutant was prepared in 20 mM MES, 200 mM KCl, 20 mM $CaCl_2$, pH 6.8. An excess of calcium was always used in the buffer to avoid exchange of Ca^{2+} from the binding sites with any free Ln^{3+} ions present in the tag solution.

NMR Measurements

All NMR experiments were performed at 298 K. ^{15}N labeled N60DCaM (0.4 mM) was titrated to $(Ca_2)_N(Ca_2)_C$-CaM and $(CaLn)_N(Ca_2)_C$-CaM (Ln^{3+} = Tm^{3+}, Tb^{3+} and Dy^{3+}) by addition of small amounts of calcium(II) and subsequently lanthanide(III) solutions. The titrations were performed by following the ^1H-^{15}N HSQC spectra at 700 MHz as previously reported.[15] ^1H-^{15}N IPAP HSQC spectra were also acquired to obtain the rdc values.

^1H-^{15}N HSQC and IPAP-HSQC spectra of Ln^{3+}-CLaNP-5 Ca_4CaM (Ln^{3+} = Lu^{3+}, Yb^{3+} and Tm^{3+}) were acquired at 298 K and 700 MHz. HNCO,[21] HNCA,[22] CBCACONH[21] and HNCACB[23] experiments at 500 MHz were performed on ^{15}N and ^{13}C labeled H107C/N111C CaM tagged with Ln^{3+}-CLaNP-5 (Ln^{3+}=Lu^{3+} and Yb^{3+}) to obtain the backbone assignment. The backbone resonance signals of Tm^{3+}-CLaNP-5 Ca_4CaM were assigned based on the assigned ^1H-^{15}N HSQC spectra of the diamagnetic Lu^{3+} form and of the paramagnetic Yb^{3+} form.

Pcs data were obtained from the difference in ^1H chemical shift between corresponding nuclei in the paramagnetic and diamagnetic CaM derivatives. Rdc data were obtained as the difference in the doublet splitting in the indirect ^{15}N dimension in ^1H-^{15}N IPAP-HSQC spectra [(15)] between the paramagnetic form and the diamagnetic form.

Maximum occurrence (MO) calculation of CaM conformations

The pcs values measured for the domain where the paramagnetic metal is located were used to calculate the magnetic susceptibility anisotropy tensors of the different metals. For the N60D $(CaLn)_N(Ca_2)_C$-CaM samples, the program FANTASIAN[24] was used to determine the anisotropy tensors. For the Ln^{3+}-CLaNP-5 Ca_4CaM samples the programs FANTASIAN and PARAMAGNETICCYANA-2.1[25;26] were used to determine the anisotropy tensors and

the position of the metal ions with respect to the backbone of the C-terminal domain. These tensors were then fixed in all subsequent calculations.

The program for the calculation of the MO of any given conformation[11] was modified to incorporate paramagnetic restraints arising from metal ions located in both protein domains. In this way, pcs and rdc measured for the C-terminal domain when the paramagnetic metal is located in the N-terminal domain could be analyzed together with pcs and rdc measured for the N-terminal domain when the metal is located in the tagged C-terminal domain. More details are reported in the Supporting Information S1 .

A total of 400 conformations with different inter-domain orientations were obtained through the program RANCH.[11;27] The MO values of each conformation were calculated from the paramagnetic restraints (pcs and rdc) obtained for the different lanthanides located in either the N- or the C-terminal domain. Taking each conformation as a starting point, a simulated annealing minimization was performed to generate an ensemble with a maximum of 15 other conformations which, together with the starting conformation, provides the best fit of the experimental data. Such fit was performed by minimizing a target function (TF) defined as the sum of the squared difference between the values obtained from the weighted average of pcs and rdc calculated for all conformations of the ensemble and the corresponding experimental data (see Eq. S2). The weight of the starting conformation was fixed in the minimization, and several calculations were repeated by changing the weight of such conformation. The MO value of each conformation was calculated as the weight for which the TF is 10% larger than the minimum value. More details are reported in the Supporting Information S1.

Results and discussion

Synthetic Tests

Synthetic tests were performed by simulating the conformational heterogeneity of a two-domain protein like calmodulin. This was done by generating a large number (50000) of protein conformations using a Gaussian probability distribution around one selected conformation. Pcs and rdc data were simulated from the average of the rdc and pcs values

obtained for the different conformations. Pcs and rdc data were calculated for four or five ions with different anisotropy tensors all located in one domain or distributed between the two domains.

These simulated data where then used to obtain the MO of the conformation at the center of the Gaussian distribution and of conformations with the orientation of the C-terminal domain described by Euler angles whose values were varied by ±50° (standard deviation of the distribution) from the central conformation. The calculations indicated that the conformation at the center of the Gaussian distribution can be better identified through the MO values when the metals are distributed in two domains rather than being all in a single domain. In fact, the difference in the MO between the central conformation and the conformations at one standard deviation increases when pcs and rdc data are referred to metals distributed in the two domains, rather than being all located in the same domain. More details are shown in the Supporting Information S2. This finding is not obvious as it has been proven that the amount of information for the characterization of the interdomain mobility is larger when rdc arising from metals with different anisotropy tensors without any main direction in common are all located in the same domain, with respect to the case of having them distributed in the two domains, in the absence of experimental errors and without using pcs.[28]

N60D $(CaLn)_N(Ca_2)_C$-CaM

Pcs and rdc for $(CaLn)_N(Ca_2)_C$-CaM (Ln^{3+} = Tm^{3+}, Tb^{3+} and Dy^{3+}) were measured in buffer 20 mM MES, 200 mM KCl and pH 6.8, the same used for the CLaNP-5 Tagged CaM samples. Some differences can be appreciated from those previously measured in 400 mM KCl, pH 6.5.[12]

The structure of the CaM domains in solution was fixed to the coordinates deposited in PDB 1J7O and 1J7P.[29] These structures were chosen because refined with an extensive use of rdc derived by external orienting media. The position of the lanthanide ions in the N60D $(CaLn)_N(Ca_2)_C$-CaM samples was fixed to the coordinates of the calcium ion in the second binding loop of the N-terminal domain.

The best fit of the pcs of the N-terminal domain amide protons to the protein structure provided the magnetic susceptibility anisotropy tensors reported in Table 1. They are in good

agreement with those previously obtained.[5;12] The quality of the fit is good as shown in Figure 1A.

The best fit of the rdc of the C-terminal domain amide protons to the protein structure provided the anisotropy tensors reported in Table 1. As previously found, they are sizably smaller that those obtained from the pcs of the N-terminal domain nuclei. The good quality of the fits, shown in Figure 1B, however indicate that the data are in good agreement with the protein structure, which thus moves as a rigid body, so that the obtained tensors are averages of the magnetic susceptibility anisotropy tensors positioned in the N-terminal domain as seen from a nucleus in the C-terminal domain.

CLaNP-5 Tagged CaM

Paramagnetic ions were placed in the C-terminal domain of CaM using the CLaNP-5 tag.[18] The mutation H107C/N111C was performed in order to allow the tag to be attached to the protein through sulfur bonds. The residues to be mutated were chosen i) positioned on one helix (the second of the C-terminal domain) in order to provide rigidity to the CLaNP-5 tag; ii) so that the cysteine side chains are exposed on the surface of the structure with the C^{β} atoms pointing away and the C^{α} atoms not closer than 6 Å and nor farther than 10 Å from one another; iii) in order to attach the tag in a position far enough from the N-terminal domain to avoid steric clashes that may affect the conformational heterogeneity of the protein.

Lu^{3+}-CLaNP-5 was used as the diamagnetic reference. The ^{1}H-^{15}N HSQC spectrum of Lu^{3+}-CLaNP-5 H107C/N111C CaM is similar to that of Ca_4CaM with differences limited to the residues in close proximity to ClaNP-5, indicating that the protein structure is maintained after binding of the tag (see Figure S4). Both Yb^{3+} and Tm^{3+} CLaNP-5 induced positive paramagnetic shifts, which in the Tm^{3+} form are much larger than in the Yb^{3+} form (see Figure 2), due to the larger magnetic susceptibility anisotropy of Tm^{3+}.

The program PARAMAGNETICCYANA-2.1 was first used to determine the position of the metal ions with respect to the C-terminal domain structure (PDB 1J7O[29]) using typical values for the magnetic susceptibility anisotropies and the observed pcs measured in the presence of Tm^{3+} or Yb^{3+} bound to the tag. The magnetic susceptibility anisotropy values were then refined using the program FANTASIAN through the best fit of the pcs to the

C-terminal domain structure and the relative calculated position of the lanthanides. The two programs were cycled iteratively until convergence of both the metal position and the susceptibility anisotropy tensors was reached.

The observed pcs values fit very well versus the calculated data (Figure 3A). The resulting axial and rhombic components of the magnetic susceptibility anisotropy tensors as well as the Euler angles providing the orientation of the tensors are reported in Table 1. These results are in agreement with the values reported in Keizers et al.[18] The *z*-axes of the Yb^{3+} and Tm^{3+} tensors are parallel, and the *x*- and *y*-axes of the two metals experience difference in the orientation of only 21° (Fig. 4). The calculations show that the metals are located at about the same distance from the protein backbone as in Keizers et al.[18] The positions of Yb^{3+} and Tm^{3+} are in fact similar for the two metals and at distances of 8.3 Å and 6.5 Å from the C^{α} atoms of residues Cys-107 and Cys-111, respectively. The magnitude of the calculated anisotropies and the correct definition of the lanthanide position indicates that the CLaNP-5 probe binds Ca_4CaM rigidly.

The rigidity of the tag is confirmed by the rdc values measured for the C-terminal domain amide protons. These rdc, when fitted to the domain structure, provide a nice agreement with the values calculated from the best fit tensors. This is clear in particular for the Yb^{3+} sample, the C-terminal domain nuclei of which are less affected by paramagnetic line broadening due to the smaller susceptibility tensor of Yb^{3+} than that of Tm^{3+}, which causes a smaller Curie relaxation (Figure 3B). The best fit anisotropy tensors ($\Delta\chi_{ax}= 8.1\times10^{-32}\,m^3$, $\Delta\chi_{rh}= -2.5\times10^{-32}\,m^3$) are actually very similar to those calculated from the pcs, indicating that no (or very modest) reduction due to motional averaging occurs. When the measured rdc are compared with the rdc values calculated from the tensor derived from pcs, a good agreement is indeed observed (Figure 3B), the differences between calculated and observed data being all within 2 Hz.

Due to the much smaller magnetic susceptibility anisotropy of Yb^{3+} with respect to Tm^{3+}, a reliable set of pcs and rdc for the N-terminal nuclei could be observed only for the Tm^{3+} derivative. The rdc, quite reduced with respect to those measured for the C-terminal domain, can be described by a single average tensor. The latter was obtained by fitting the rdc to the N-terminal domain CaM structure. As expected, the calculated average tensor is

sizably smaller than that obtained from the C-terminal domain pcs data (Table 1) as a consequence of extensive orientation averaging. The fit of the observed rdc versus the calculated values is shown in Fig. 3C. Pcs of the N-terminal domain ^{N}H nuclei were also collected, and they are reported in Fig 3D.

MO analysis

Calculations of the MO values for 400 CaM conformations randomly generated have been performed using the derived magnetic susceptibility anisotropy tensors and the pcs and rdc data observed for the domain without the paramagnetic metal. MO values are obtained from the largest weight that each of the 400 conformations can have when included in any possible ensemble with other 15 conformations with different weights. This ensemble was found as the family of structures in best agreement with the experimental data by minimizing the target function (TF), defined as a measure of the disagreement from the experimental data of the pcs and rdc values calculated according to the ensemble itself (see Materials and Methods and Supporting Information S1 for further details). It was checked that increasing the number of conformations above 15 does not decrease the TF, so that such number of conformations was chosen for the calculations. During the minimization, the weight of the fixed conformation (one of the 400 randomly generated conformations) was changed. The MO of such conformation was set to the largest weight for which the TF is smaller than a given threshold. The latter was defined 10% larger than the lowest possible TF value.

The results obtained for the Tb^{3+}, Tm^{3+} and Dy^{3+} ions positioned in the N-terminal domain provide the map of MO values shown in Fig 5A. The position of the C-terminal domain of CaM is indicated by an orientation tensor centered in the center of mass of the C-terminal domain, color-coded with respect to the MO of the corresponding conformation from blue (lower than 5%) to red (greater than 30%). Different orientations of the tensor reflect different orientations of the CaM C-terminal domain with respect to the N-terminal domain. The minimum for the TF was calculated by generating structural ensembles without any fixed conformation, and resulted equal to 0.203, so that a threshold of 0.223 was fixed. The overall distribution of the MO values is indeed relatively similar to that previously calculated for data acquired with a higher salt concentration in solution and with inclusion of SAXS restraints (Figure 3A of Bertini et al.[11]). As already seen, the conformations having the

C-terminal domain in the lower right quadrant of the frame have in general low MO, while the conformations with the highest MO are clustered in the central part of the distribution, corresponding to relatively but not fully elongated conformations.

The set of rdc data acquired for the N-terminal domain when the Tm^{3+} tag is placed in the C-terminal domain of the CaM mutant was then added to the previous data, and MO calculations were repeated for the same 400 conformations. In this case, a minimum for the TF was calculated equal to 0.224, so that a threshold of 0.246 was fixed. The number of conformations with TF smaller than the defined threshold when their weight was 0.1 increased from 66 to 248 (Figure 6). As expected, this indicates that the new set of data is effective in decreasing selectively the MO of conformations much less sampled or even not actually sampled by the system. The MO values are shown in Fig 5B. Figure 7 shows the *TF* values for all the conformations as a function of their weight. The substantial differences in the weight at which the *TF* value start increasing result in markedly different MO.

Figure 8 shows the conformations with a difference in the MO values upon inclusion of the last set of restraints larger than 0.1. The corresponding orientation tensors are color-coded with respect to the MO difference from blue (difference in MO of 0.10) to red (difference in MO of 0.25). The figure shows that the effect is distributed along all the conformational space, for all relative positions and orientations of the C-terminal domain.

The present calculations, as well as the simulations performed with synthetic data, show that although rdc arising from up to 5 metals located in different domains are not fully independent,[28] differently from when they are all located in the same domain, they are quite informative for the determination of the MO of the different conformations when coupled with pcs and considering that data are effected by experimental errors. The information on the interdomain conformational variability that is contained in the rdc arising from different metals placed in the same domain is in fact larger than that contained in the rdc arising from the same number of metals but distributed between the two domains.[28] In the latter case, however, relationships are present among the average rdc-derived tensors and the pcs-derived magnetic susceptibility anisotropy tensors[28] that can profitably assess the consistency of the data. We have shown in the simulations reported in the Supporting Information that the use of pcs in the calculations may however compensate the smaller

information content of rdc on the protein conformational distribution, and may make the distribution of the metals in both protein domains preferable. Therefore, the use pcs and rdc data arising from metals placed in both protein domains i) permits an internal check of the quality of the data from the overall agreement of all sets of the experimental rdc with those calculated from the best-fit conformational ensembles and ii) provides a better discrimination among the different protein conformations depending on the calculated MO values.

Conclusions

We have shown that the simultaneous use of paramagnetism-based restraints arising from paramagnetic metal ions located in different domains of proteins experiencing interdomain mobility is quite informative for the determination of the maximum occurrence of any conformation. The increase in the information content provided by locating paramagnetic ions in different domains is larger than obtained by placing the same number of metal ions in the same site especially because of the quite different orientations of the magnetic susceptibility anisotropy tensors. In the case of CaM we have shown that the maximum occurrence of several conformations is quite reduced by the addition of pcs and rdc arising from the presence of a single metal ion rigidly attached to the C-terminal domain with respect to the values calculated using three metal ions placed in the N-terminal domain of the protein. As a result, the conformations likely experienced by the protein can be more accurately mapped.

Reference List

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

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

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

(4) Kay, L. E. *J.Magn.Reson.* **2005**, *173*, 193-207.

(5) 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.

(6) Xu, X.; Reinle, W.; Hannemann, F.; Konarev, P. V.; Svergun, D. I.; Bernhardt, R.; Ubbink, M. *J.Am.Chem.Soc.* **2008**, *130*, 6395-6403.

(7) Iwahara, J.; Clore, G. M. *Nature* **2006**, *440*, 1227-1230.

(8) Tang, C.; Schwieters, C. D.; Clore, G. M. *Nature* **2007**, *449* , 1078-1082.

(9) Volkov, A. N.; Worrall, J. A. R.; Holtzmann, E.; Ubbink, M. *Proc.Natl.Acad.Sci.USA* **2006**, *103*, 18945-18950.

(10) Bashir, Q.; Volkov, A. N.; Ullmann, G. M.; Ubbink, M. *J.Am.Chem.Soc.* **2010**, *132*, 241-247.

(11) Bertini, I.; Giachetti, A.; Luchinat, C.; Parigi, G.; Petoukhov, M. V.; Pierattelli, R.; Ravera, E.; Svergun, D. I. *J.Am.Chem.Soc.* **2010**, *132*, 13553-13558.

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

(13) Longinetti, M.; Luchinat, C.; Parigi, G.; Sgheri, L. *Inv.Probl.* **2006**, *22*, 1485-1502.

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

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

(16) Bertini, I.; Kursula, P.; Luchinat, C.; Parigi, G.; Vahokoski, J.; Willmans, M.; Yuan, J. *J.Am.Chem.Soc.* **2009**, *131*, 5134-5144.

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

(18) Keizers, P. H. J.; Saragliadis, A.; Hiruma, Y.; Overhand, M.; Ubbink, M. *J.Am.Chem.Soc.* **2008**, *130*, 14802-14812.

(19) Keizers, P. H.; Desreux, J. F.; Overhand, M.; Ubbink, M. *J.Am.Chem.Soc.* **2007**, *129*, 9292-9293.

(20) Bertini, I.; Luchinat, C.; Parigi, G.; Pierattelli, R. *Dalton Trans.* **2008**, *2008*, 3782-3790.

(21) Muhandiram, D. R.; Kay, L. E. *J.Magn.Reson.Ser.B* **1994**, *103*, 203-216.

(22) Bax, A.; Ikura, M. *J.Biomol.NMR* **1991**, *1*, 99-104.

(23) Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A. *J.Magn.Reson.* **1990**, *89*, 496-514.

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

(25) Banci, L.; Bertini, I.; Bren, K. L.; Cremonini, M. A.; Gray, H. B.; Luchinat, C.; Turano, P. *J.Biol.Inorg.Chem.* **1996**, *1*, 117-126.

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

(27) Bernado, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. *J.Am.Chem.Soc.* **2007**, *129*, 5656-5664.

(28) Sgheri, L. *Inv.Probl.* **2010**, *26*, 115021-115021-12.

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

Table 1. Magnetic susceptibility anisotropies and average tensors of the different lanthanides in the second calcium binding loop of the N-terminal domain of CaM and in the ClaNP-5 CaM. The Euler angles are in the ZYZ convention.

	$\Delta\chi_{ax}$ (10^{-32}m^3)	$\Delta\chi_{rh}$ (10^{-32}m^3)	Euler angles (P, T, O)		
Metals in the N-terminal domain – from N-terminal domain pcs					
Tb	35.6	-16.5	1.767[a]	-0.883[a]	0.709[a]
Tm	30.7	-8.5	0.476[a]	-0.510[a]	1.812[a]
Dy	35.5	-12.9	1.317[a]	-0.721[a]	0.312[a]
Metals in the C-terminal domain – from C-terminal domain pcs					
Yb	9.7	-2.6	0.885[b]	-1.715[b]	1.822[b]
Tm	56.3	-6.7	0.834[b]	-1.360[b]	1.882[b]
	$\overline{\Delta\chi}_{ax}$ (10^{-32}m^3)	$\overline{\Delta\chi}_{rh}$ (10^{-32}m^3)	Euler angles (P, T, O)		
Metals in the N-terminal domain – from C-terminal domain rdc					
Tb	3.1	2.9	-2.625[b]	-0.042[b]	0.041[b]
Tm	2.0	1.0	-0.324[b]	0.568[b]	-0.160[b]
Dy	2.7	-1.6	1.703[b]	0.049[b]	0.038[b]
Metals in the C-terminal domain – from N-terminal domain rdc					
Tm	3.0	2.9	3.106[a]	0.301[a]	-0.336[a]

[a]with respect to structure 1J7O; [b]with respect to structure 1J7P

Figure 1. (A) Observed versus calculated values of pcs of N-terminal domain nuclei for the terbium(III), thulium(III) and dysprosium(III) ions substituted in the second binding site of CaM N-terminal domain. (B) Observed versus calculated values of rdc of C-terminal domain HN for the terbium(III), thulium(III) and dysprosium(III) ions substituted in the second binding site of CaM N-terminal domain.

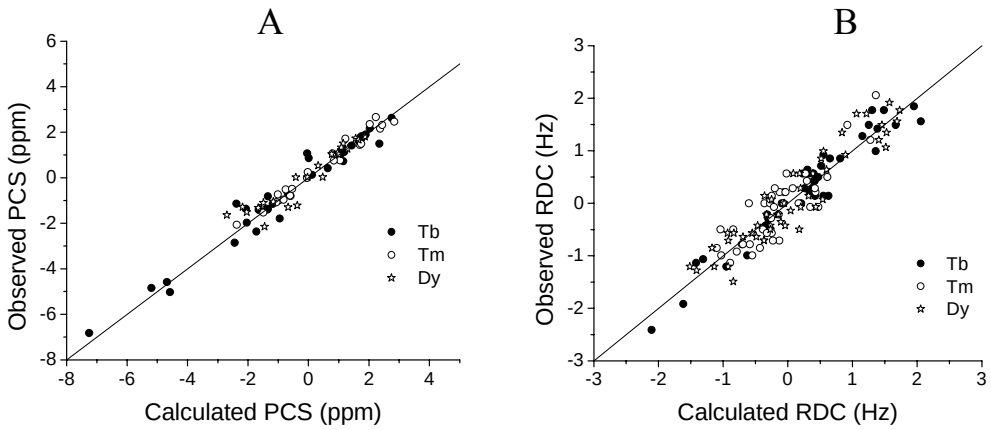


Figure 2 Superposition of a region of the ^{1}H-^{15}N HSQC spectra of Lu^{3+} (red), Yb^{3+} (green) and Tm^{3+} (blue) ClaNP-5 CaM

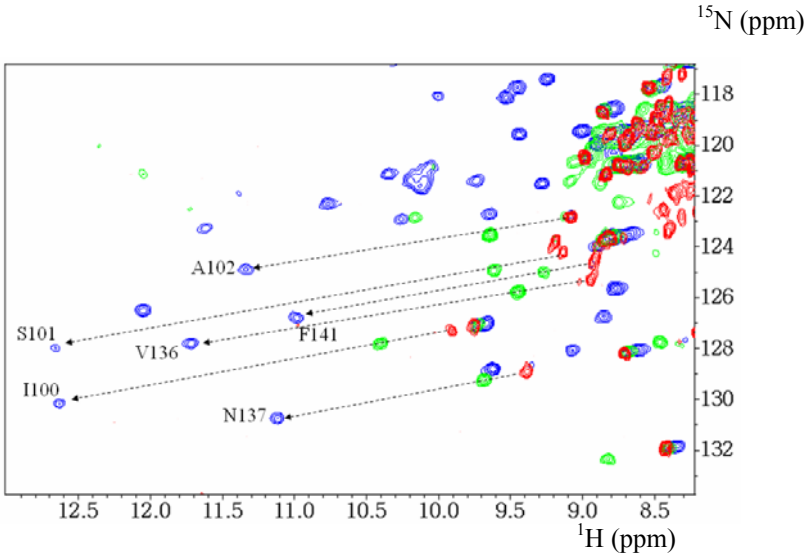


Figure 3 (A) Observed versus calculated pcs of C-terminal domain nuclei for the Yb^{3+} (●) and Tm^{3+} (○) ClaNP-5 Ca_4CaM. (B) Observed versus calculated rdc of C-terminal domain HN for the Yb^{3+}-ClaNP-5 Ca_4CaM. The solid symbols (●) indicate the values calculated from the best fit parameters ($\Delta\chi_{ax}$= 8.1×10^{-32} m^3, $\Delta\chi_{rh}$= -2.5×10^{-32} m^3), the open symbols (○) indicate the values calculated using the pcs-derived tensor. (C) Observed versus calculated rdc of N-terminal domain HN for the Tm^{3+} ClaNP-5 Ca_4CaM. (D) Observed pcs of the C-terminal domain for Tm^{3+}-ClaNP-5 Ca_4CaM.

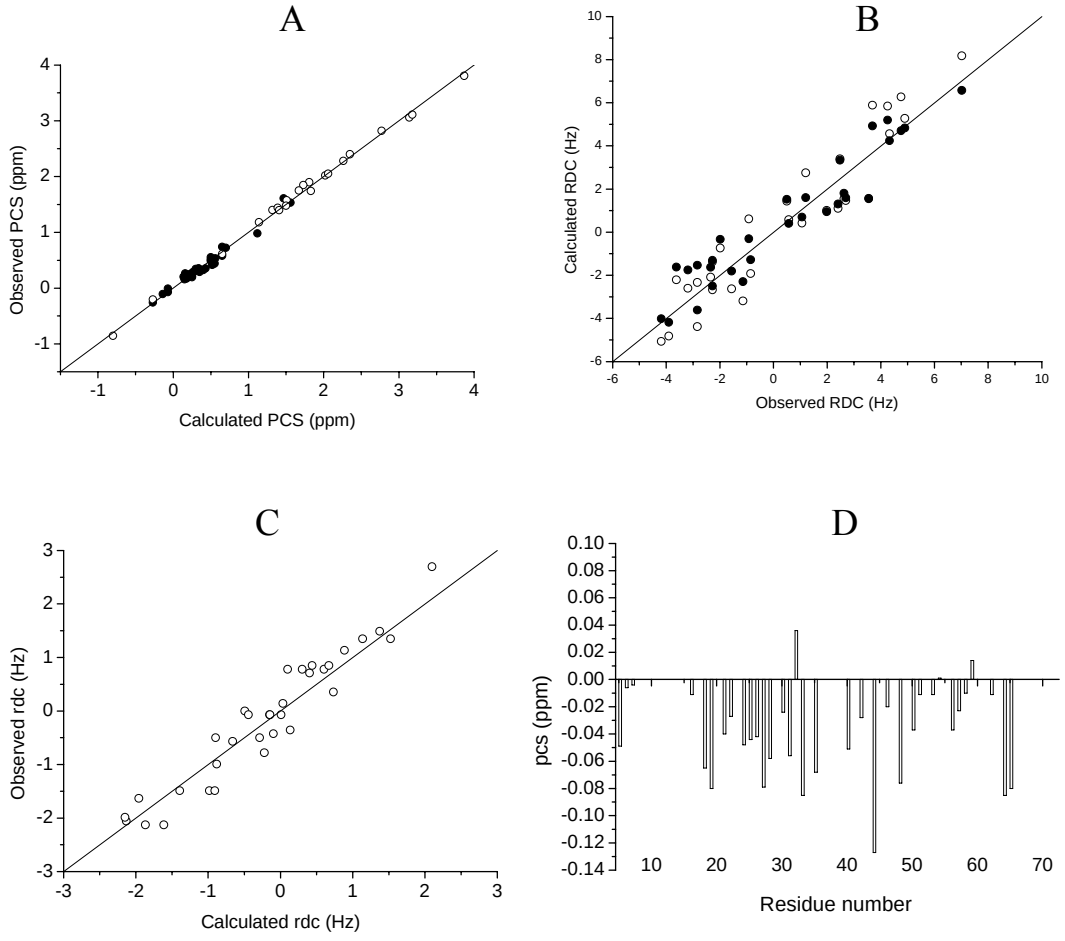


Figure 4. The lanthanide ions (in blue) are placed at a distance of 8.3 Å and 6.5Å from the C^{α} atoms of residues Cys-107 and Cys-111. The orientations of the magnetic susceptibility anisotropy tensors are shown for the Yb^{3+} (magenta) and Tm^{3+} (brown) metals. The *z* axes of the anisotropy tensors of the two metals are essentially coinciding; the angle between the *x* (and *y*) axes of the anisotropy tensors of the two metals is 21.4°.

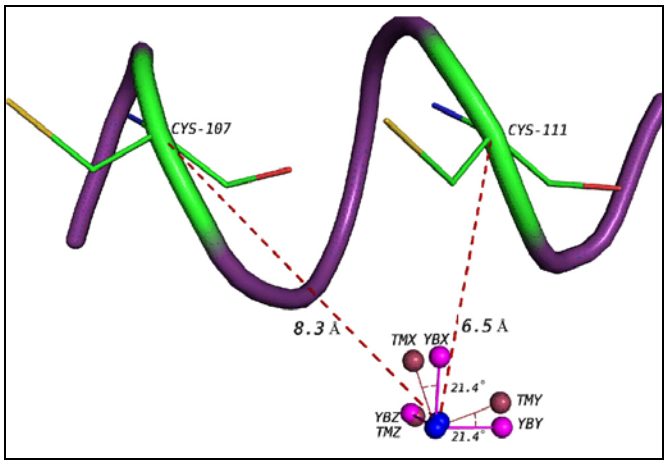


Figure 5. Arbitrary orientation tensors centered in the center of mass of the C-terminal domain, color-coded with respect to the MO of the corresponding conformation from blue (lower than 5%) to red (greater than 30%) for 400 structures generated randomly with RANCH. Panel A shows the results obtained from pcs and rdc arising with metals in the N-terminal domain; panel B shows the results obtained when pcs and rdc of Tm^{3+}-ClaNP-5 Ca_4CaM are also included.

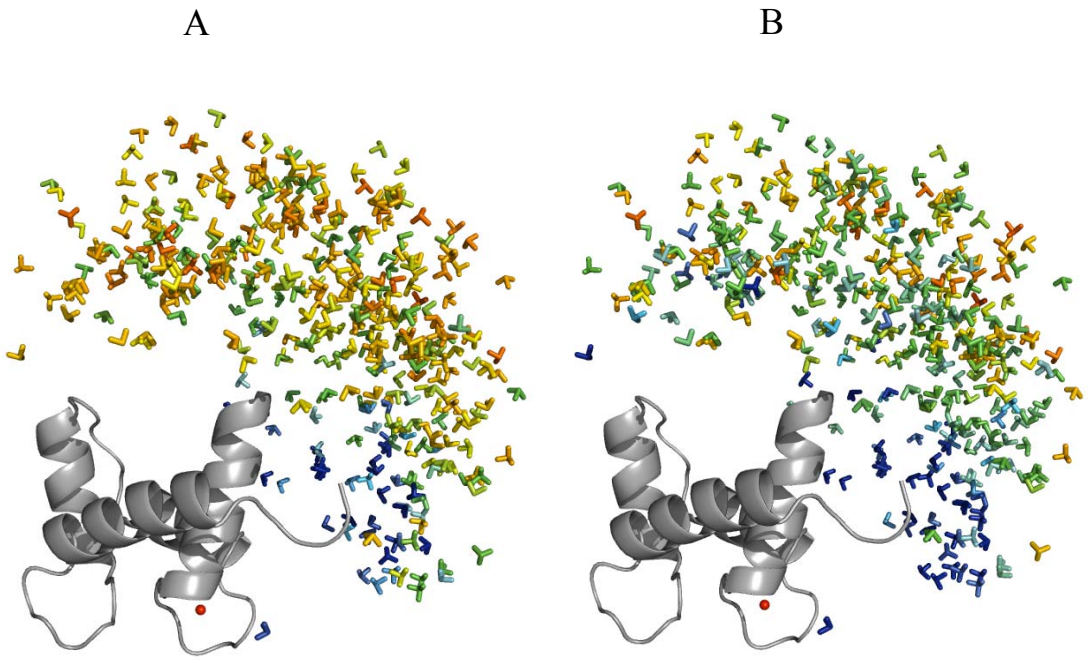


Figure 6. Number of conformations as a function of the MO values calculated from the pcs and rdc arising with metals in the N-terminal domain (3+0 case) and by including pcs and rdc of Tm^{3+}-ClaNP-5 Ca_4CaM (3+1 case).

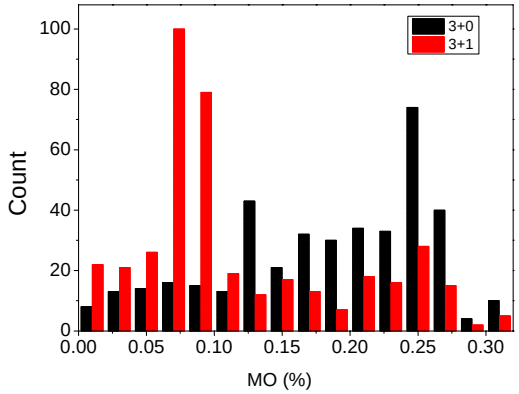


Figure 7. TF (describing the best agreement between any conformational ensemble and the experimental data) as a function of the weight of the conformation for which the MO is calculated. Each curve corresponds to a different conformation. The MO is defined by the intersection between the TF curve and the threshold chosen 10% larger than the smallest TF. Panel A shows the results obtained from pcs and rdc arising with metals in the N-terminal domain; panel B shows the results obtained when pcs and rdc of Tm^{3+}-ClaNP-5 Ca_4CaM are also included.

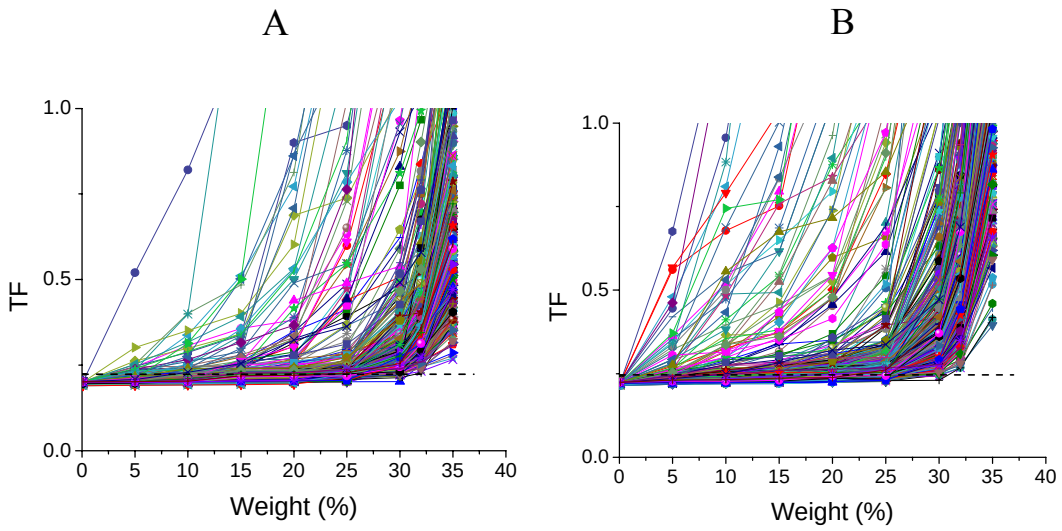
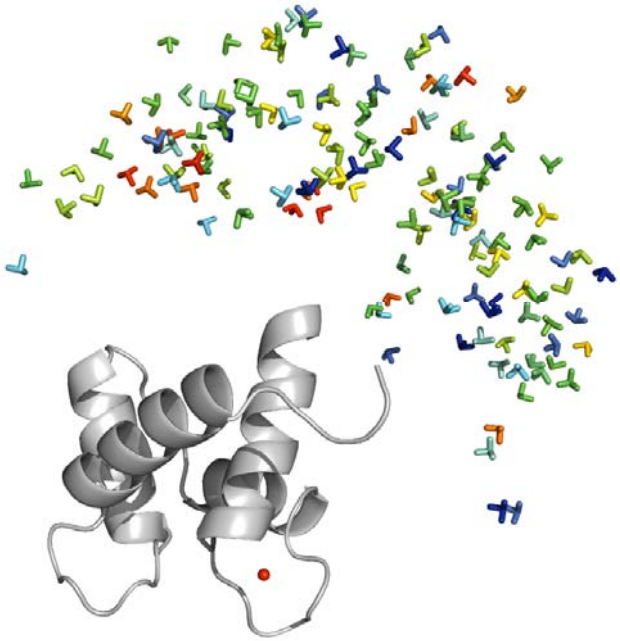


Figure 8. Arbitrary orientation tensors centered in the center of mass of the C-terminal domain, color-coded with respect to the difference in MO values upon inclusion of pcs and rdc of Tm^{3+}-ClaNP-5 CaM. The colors change from blue (difference in MO of 0.1) to red (difference in MO of 0.25) for the structures with MO differences larger than 0.1.



Supporting information

S1: MO calculation

The MO (maximum occurrence) values for hundreds of conformations of CaM have been computed using the paramagnetic restraints pcs and rdc obtained from metal ions positioned in two different metal binding sites located in protein domains which are mobile with respect to the one another. The calculations for the different conformations were performed simultaneously on a grid system.

A subroutine was implemented into the MO program to calculate the contribution to the target function (TF) arising from violations in the pcs and rdc values of nuclei located in the moving domain with respect to the reference frame fixed on the metal containing domain. Starting from a reference protein conformation, the relative orientation of the C-terminal domain with respect to the reference frame fixed on the N-terminal domain was calculated as a clockwise rotation; the opposite (anticlockwise) rotation was applied to determine the inter-domain orientation when the reference frame is fixed on the C-terminal domain. Similarly, a translation with the same amplitude but in the opposite direction was performed to compute the relative position of the domains.

In the present study, when the metal ion is coordinated to the N-terminal domain, the different conformations of CaM are obtained by moving the C-terminal domain from a position *P(x,y,z)* to a new position *P'* through a rotation *R* and a translation *t*. The new position *P'* of the C-terminal domain was computed as *R**(*P*-*t*). On the other hand, when the metal ion is fixed to the C-terminal domain, the same conformation of CaM is obtained by moving the N-terminal domain to the corresponding relative position *P'* computed as (R^{-1}**P*) +*t*. The R rotation matrix was obtained from the Euler angles in the ZYZ convention:

$$R = \begin{bmatrix} c_1c_2c_3 - s_1s_3 & -c_2s_3c_1 - c_3s_1 & c_1s_2 \\ c_1s_3 + c_3c_2s_1 & c_1c_3 - c_2s_1s_3 & s_2s_1 \\ -c_3s_2 & s_3s_2 & c_2 \end{bmatrix} \qquad [S1]$$

where c and s mean cosine and sine, and the subscripted numbers 1,2,3 indicate the three Euler angles (P,O,T).

The protocol used to calculate the MO values was the following:

1. A set of 400 fixed conformations representing the conformational space sampled by CaM was generated using the program RANCH [(1)]. The following calculations

were then performed for each conformation.

2. A weight smaller than 100% was fixed for the selected conformation. Other conformations with different weights (the sum of all weights being constrained to 1) were then added one by one, and each time a simulated annealing minimization was performed, in order to obtain the best possible agreement with the experimental data. A maximum number of 15 conformations was used to best fit the experimental observables. The TF to be minimized was defined as

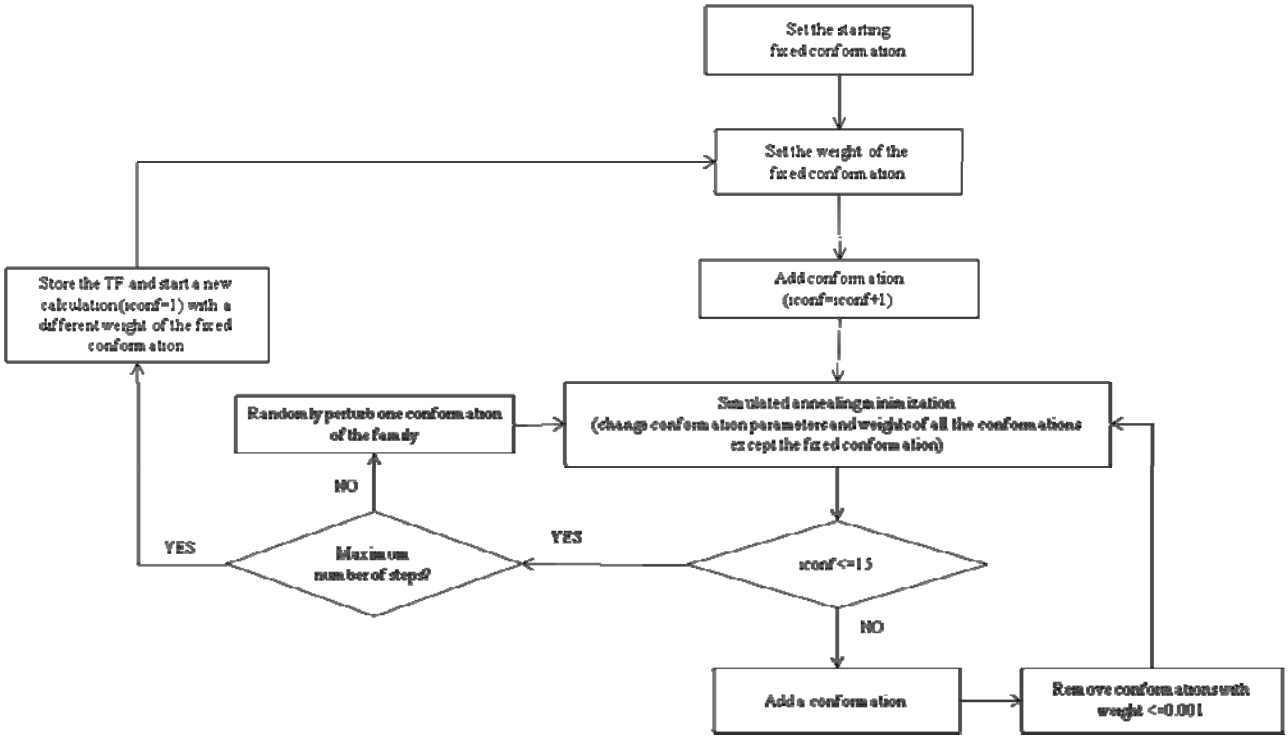
$$TF(w_0) = \min_{t_0,(w_i,t_i,R_i)} \sum_j \left| \tilde{\delta}_j - \left(w_0 \delta_j(t_0, R_0) + \sum_{i=1}^{N} w_i \delta_j(t_i, R_i) \right) \right|^2 \qquad \text{[S2]}$$

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 conformation with orientation R_0 and translation vector t_0, 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 with weight w_i, position t_i, and orientation R_i. During the minimization, conformations with weight less than 10^{-3} were removed, in order to ensure convergence more rapidly.

3. The weight of the selected conformation was changed and step 2 was repeated (using the ensembles calculated with different weights as starting points for the minimization).

4. The MO of each selected conformation was calculated as the weight at which the TF is 10% larger than the minimum value achieved at low weights.

Figure S1.1 Flow chart of MO calculations



S2: Synthetic tests

A simulation was performed by generating a family of 50000 conformations with a Gaussian distribution around one selected conformation, shown in Fig. S2.1, corresponding to the PDB 1CLL conformation. The orientation of the C-terminal domain with respect to that of the N-terminal domain of any conformation of the family is defined by 3 Euler angles. Pcs and rdc data corresponding to metals (with their anisotropy tensors) positioned in either the N or the C-terminal domains were then calculated according to the generated family of conformations.

Calculations were then performed using different sets of pcs and rdc data to check the sensitivity of MO to the different probability of the conformations within the family. The MO values of the conformation at the centre of the Gaussian distribution and of other 8 conformations defined by changing each of the 3 Euler angles of ±50° (standard deviation of the distribution) from the central conformation were calculated.

A signal to noise is defined as the ratio between the MO of the conformation at the Gaussian centre and the average of the MO values of the other 8 conformations:

$$\frac{\text{Signal}}{\text{Noise}} = \frac{\text{MAP of conformation at Gaussian centre}}{\text{Average MAP of the other 8 conformations}} \qquad [3]$$

A larger signal to noise indicates an increased capability of the MO values to discriminate the conformations with larger probability from those with smaller probability. One of the 8 conformations was actually excluded in the performed calculations because in such conformation the two domains crashed.

The calculations were performed both without and with including an experimental error on the simulated pcs and rdc data, using a total of 4 or 5 metal ions with different anisotropy tensors. Without errors, the signal to noise ratio was almost the same (1.26), irrespective of the distribution of the lanthanides within the two domains, if only rdc data are used in the fit. A small difference in the signal to noise ratio was observed when a Gaussian error was added to the rdc data. The best signal to noise ratio of 1.36 was obtained in the case of 3 metals in the N-terminal domain and 2 metals in the C-terminal domain (3+2 case) relative to the other distributions of the metal ions in the two domains (see Table S2.1).

The same analysis was performed including the pcs data into the calculations. MO values were obtained for the central and the other 7 conformations using pcs and

rdc data without error (Fig.S2.2) and with error (pcs' and rdc') (Fig.S2.3). The pcs' and rdc' values are defined as:

$$pcs' = pcs \times (1 + Yg_1 \times 0.01) + Yg_2 \times 0.01 \quad [4]$$

$$rdc' = rdc \times (1 + Yg_1 \times 0.1) + Yg_2 \times 0.1 \quad [5]$$

$$Yg_1 = \sqrt{-2 \times \log(Xg_1) \times \sin(\pi \times Xg_2)} \quad [6]$$

$$Yg_2 = \sqrt{-2 \times \log(Xg_3) \times \cos(\pi \times Xg_4)} \quad [7]$$

where Xg_1, Xg_2, Xg_3 and Xg_4 are four random numbers (from 0 to 1).

Table S2.1 reports the signal to noise ratio obtained from simulated data calculated with 4 or 5 metal ions differently distributed in the two protein domains. Both with and without inclusion of an experimental error, the presence of metals in different domains is definitely advantageous. The signal to noise ratios in the cases of 3+1 and 2+2 metal distributions (number of metals in N-terminal + C-terminal domains) are always higher than in the 4+0 case.

Table S2.1 Signal to noise ratio for the performed simulations

Metals in N-terminal + C-terminal domains	Signal to noise ratio			
	Rdc restraints (no error)	Rdc restraints (with error)	rdc+pcs restraints (no error)	rdc+pcs restraints (with error)
4+0	1.24	1.23	1.28	1.35
3+1	1.26	1.33	1.37	1.43
2+2	1.25	1.20	1.54	1.42
5+0	1.23	1.29	1.33	1.47
3+2	1.26	1.36	1.60	1.51

References

(1) Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. (2007), J Am Chem Soc, 129: 5656-64.

Fig. S2.1 Gaussian distribution of the conformational family (magenta) around one selected conformation (yellow)

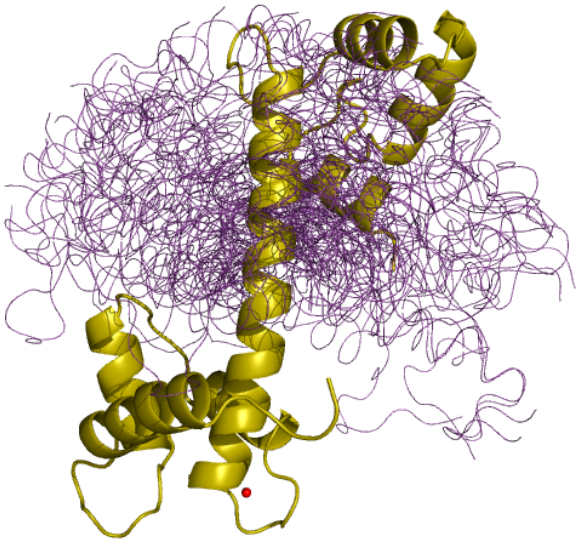
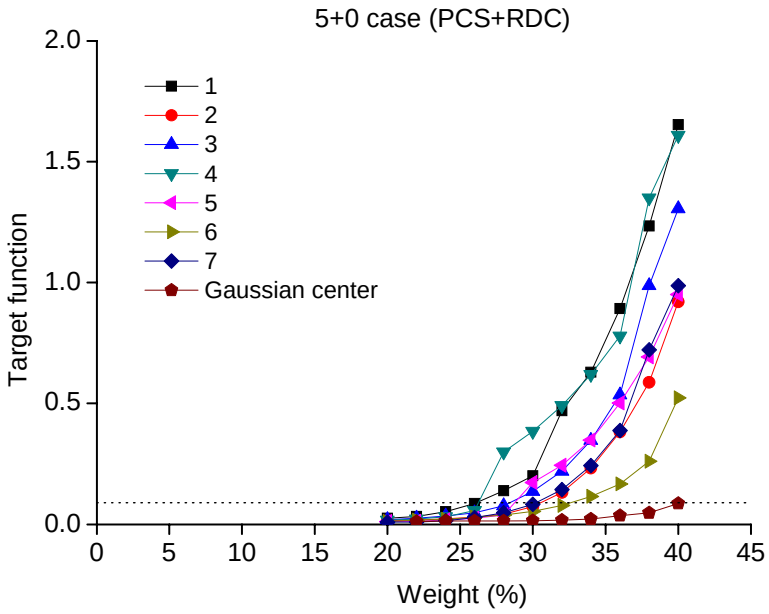


Fig. S2.2. MO analysis of the selected conformations within the simulated Gaussian distribution. Pcs and rdc data are simulated without including errors.



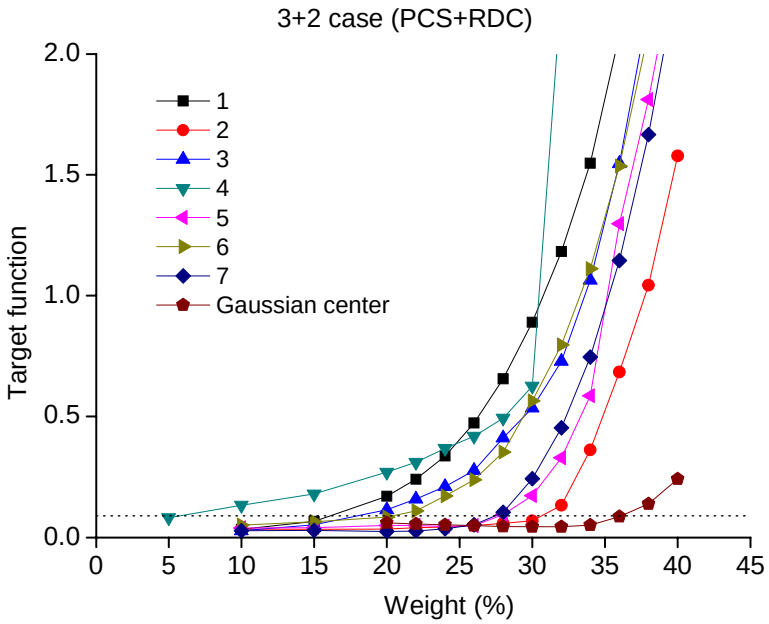
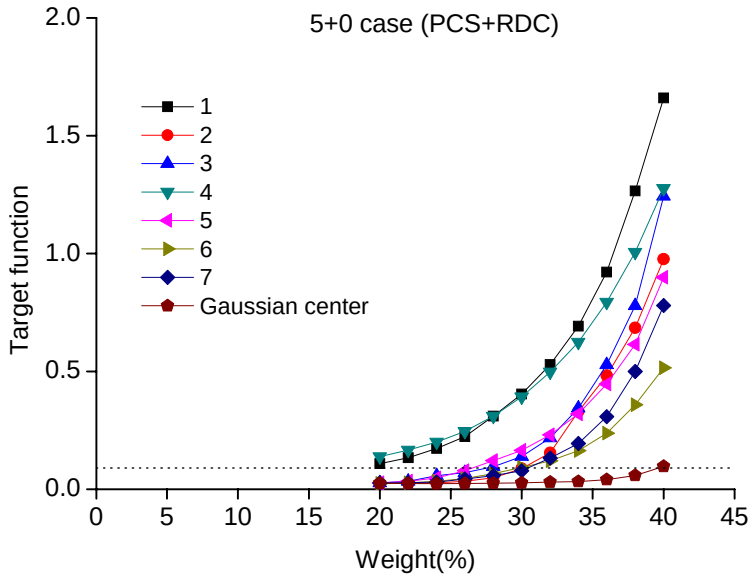
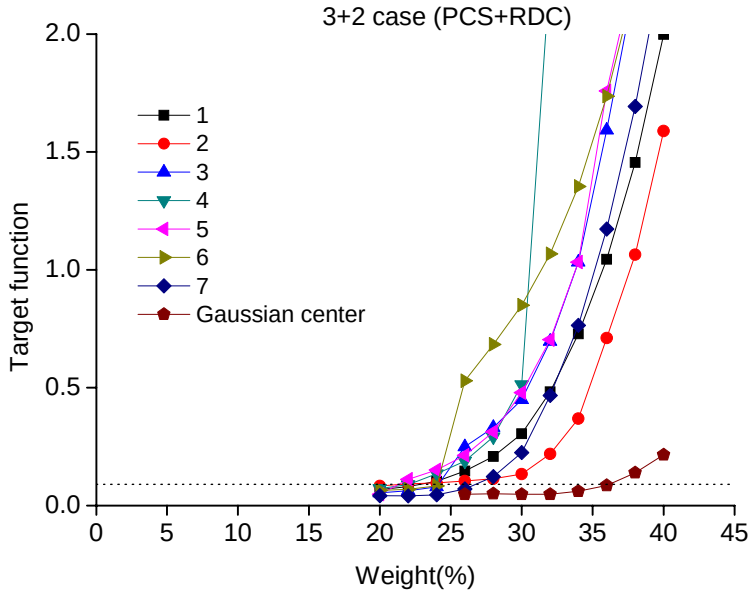


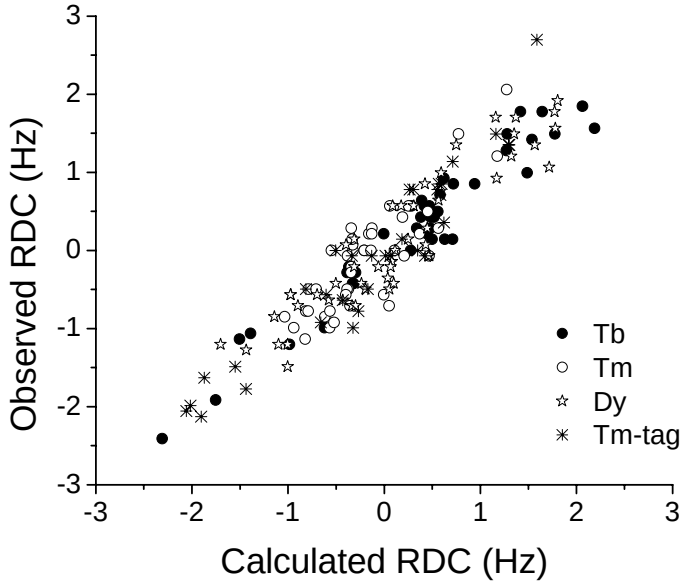
Fig. S2.3. MO analysis of the eight selected conformations within the simulated Gaussian distribution . Pcs and rdc data are simulated with addition of a Gaussian error.

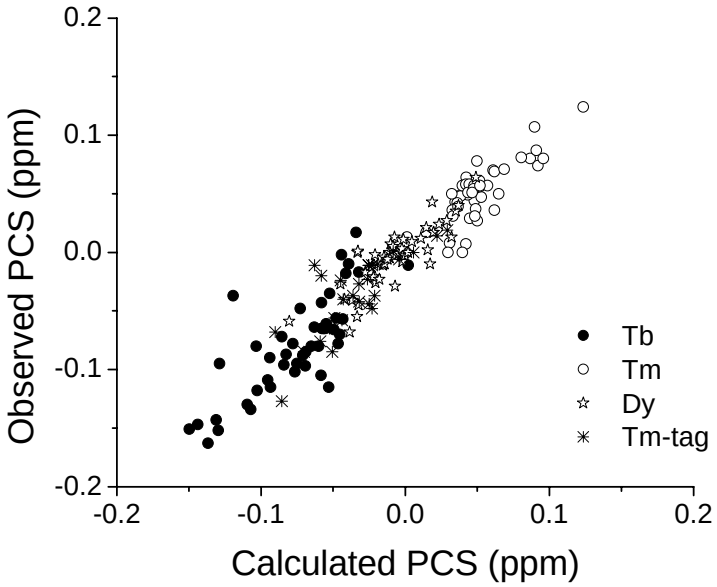




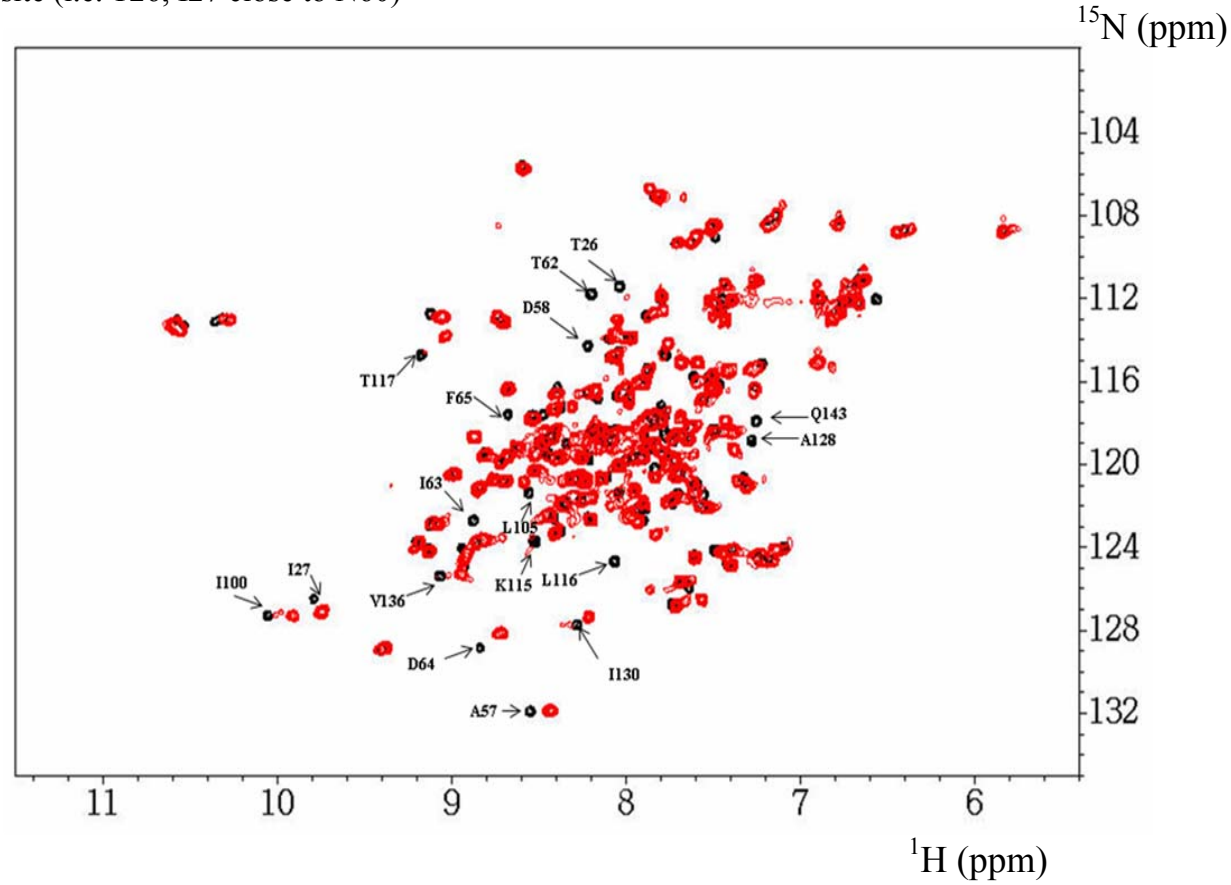
S3: Best fit graphs

Correlation between rdc (A) and pcs (B) values measured for nuclei belonging to the domain without the lanthanide ion and rdc and pcs values calculated from the ensemble of conformations providing a minimum value of TF in the MO calculations.





S4. Figure S4. ^{1}H-^{15}N HSQC spectra of Lu^{3+}-CLaNP-5 H107C/N111C CaM (red) and of N60D Ca_4CaM (black). Moving peaks are those around the mutation sites, i.e. around residues 60, 107, 111 or belonging to residues close in space to the mutation site (i.e. T26, I27 close to N60)



S5. Experimental restraints

RDC of N60D CaM				
	Residue	**Tb**	**Tm**	**Dy**
4	LEU	-	-	-0.142
5	THR	-18.447	0.098611	-17.808
6	GLU	6.953	1.064	4.683
7	GLU	3.335	-	7.024
9	ILE	5.818	2.483	3.264
10	ALA	8.585	-7.450	-9.153
11	GLU	2.199	-	-
12	PHE	-3.335	-1.419	-5.605
13	LYS+	16.177	-	12.984
14	GLU	5.818	-	-
15	ALA	-	-3.831	9.011
16	PHE	-	-	9.862
17	SER	15.254	-	-
18	LEU	-	-5.534	-
19	PHE	-0.851	-2.341	-
22	ASP	-	-9.862	-
23	GLY	-14.048	11.565	-2.909
24	ASP	-	7.946	-16.319
29	THR	-	-1.561	-0.922
30	LYS+	2.767	-	-
31	GLU	2.554	12.629	-
32	LEU	2.554	-	-
33	GLY	3.477	9.294	-
34	THR	-	10.572	-1.206
35	VAL	-	11.281	-
36	MET	-	7.450	-2.483
37	ARG+	-17.241	-	-11.281
38	SER	-10.430	-	-20.292
39	LEU	-10.501	-	-
40	GLY	-9.153	4.115	4.044
41	GLN	-14.261		4.825
44	THR	-	-1.774	1.490
45	GLU	-	-1.845	-
47	GLU	-24.407	-	-
48	LEU	-	-0.993	-7.734
49	GLN	-	1.135	
50	ASP	-	8.656	1.490
84	GLU	-	-	-0.639
85	ILE	-	0.497	-0.639
86	ARG+	-	-	-0.851
89	PHE	-	1.490	-0.710
90	ARG+	-	1.206	-1.206
91	VAL	-0.426	-	0.284

RDC of CLaNP-5-CaM			
	Residue	**Yb**	**Tm**
7	GLU	-	1.348
16	PHE	-	1.135
21	LYS+	-	0.780
22	ASP	-	0.000
23	GLY	-	0.000
24	ASP	-	0.709
25	GLY	-	-2.058
26	THR	-	-1.987
28	THR	-	-2.129
29	THR	-	-0.780
30	LYS+	-	-0.071
32	LEU	-	-0.497
33	GLY	-	0.851
35	VAL	-	-0.993
38	SER	-	-0.639
40	GLY	-	1.348
42	ASN	-	0.780
45	GLU	-	0.780
46	ALA	-	-0.071
48	LEU	-	-0.071
49	GLN	-	0.355
50	ASP	-	0.142
52	ILE	-	2.696
53	ASN	-	0.851
56	ASP	-	-0.922
57	ALA	-	-0.071
58	ASP	-	-1.490
59	GLY	-	-1.632
60	ASP	-	1.490
62	THR	-	-0.568
63	ILE	-	-1.774
64	ASP	-	-0.497
65	PHE	-	-0.071
85	ILE	-3.193	19.866
86	ARG+	2.980	3.193
87	GLU	3.264	21.994
88	ALA	3.831	22.562
89	PHE	3.831	22.562
90	ARG+	2.767	11.068
91	VAL	2.625	-14.758
92	PHE	2.412	21.143
93	ASP	2.838	21.569
94	LYS+	-1.561	-5.037

93	ASP	-2.412	2.058	-1.277
94	LYS+	-0.284	0.568	-1.206
95	ASP	1.490	0.000	0.993
96	GLY	-	-0.497	0.142
97	ASN	-1.206	0.213	-1.490
98	GLY	1.774	-0.497	1.490
99	TYR	0.993	-0.780	1.916
100	ILE	1.277	-0.780	1.064
101	SER	1.845	-0.497	1.348
102	ALA	0.426	-0.710	0.142
103	ALA	0.284	0.000	0.568
105	LEU	0.851	-0.568	0.071
106	ARG+	0.000	0.568	0.568
107	HIS+	-	-	0.568
109	MET	0.497	-	-
110	THR	0.568	-	-0.142
111	ASN	0.213	0.000	-
112	LEU	-	-	0.851
113	GLY	-	-	-0.213
114	GLU	-0.213	-	-0.568
115	LYS+	0.142	0.284	-0.355
116	LEU	0.426	0.000	-0.709
117	THR	1.774	-0.851	1.703
118	ASP	-1.064	0.426	-0.426
119	GLU	-0.993	-0.497	0.071
121	VAL	-	-0.071	-0.426
122	ASP	-1.135	0.000	-0.568
123	GLU	-	-0.780	-0.497
124	MET	0.213	-0.993	0.640
127	GLU	0.426	-1.135	1.348
128	ALA	-	0.071	-
129	ASP	-1.916	0.213	-1.206
130	ILE	0.000	0.284	-0.568
131	ASP	1.490	-0.922	1.206
132	GLY	-0.284	-0.071	-0.497
134	GLY	0.710	-0.213	0.922
135	GLN	0.851	-0.993	1.703
136	VAL	1.419	-0.851	1.774
137	ASN	1.561	-	1.561
138	TYR	0.142	-	-
140	GLU	0.568	0.284	-0.071
141	PHE	0.922	-0.071	0.000
142	VAL	0.639	-0.284	-0.213
143	GLN	0.284	0.213	-0.426
144	MET	0.142	-	-
145	MET	0.568	-0.710	-0.709

95	ASP	3.477	21.853
96	GLY	-2.270	-13.126
97	ASN	-3.902	-1.703
98	GLY	-2.483	4.683
99	TYR	4.257	0.142
100	ILE	4.896	23.839
101	SER	7.024	9.011
102	ALA	-1.987	-6.953
103	ALA	-5.037	19.937
104	GLU	0.497	-
105	LEU	-2.341	0.497
112	LEU	3.548	-
113	GLY	-1.561	20.859
114	GLU	3.547	-
115	LYS+	3.264	2.129
116	LEU	9.436	23.626
117	THR	-6.669	18.589
118	ASP	1.703	21.924
119	GLU	0.710	-6.527
120	GLU	3.547	-
121	VAL	-2.838	20.717
122	ASP	-1.206	21.782
123	GLU	3.335	22.065
124	MET	3.547	-
125	ILE	2.483	21.214
126	ARG+	-0.142	-
127	GLU	2.412	5.676
128	ALA	2.483	21.214
129	ASP	3.264	-17.808
130	ILE	1.206	3.264
131	ASP	2.483	27.529
132	GLY	-0.213	-5.889
133	ASP	-4.186	-11.849
134	GLY	1.064	12.345
135	GLN	2.696	12.487
136	VAL	4.328	20.008
137	ASN	3.193	26.819
138	TYR	8.656	21.64
139	GLU	0.355	-11.636
140	GLU	1.064	22.562
141	PHE	0.568	-9.082
142	VAL	-2.341	21.498
143	GLN	3.547	-
144	MET	1.987	3.831
145	MET	-2.270	-16.744

147	ALA	-0.426	0.142	-0.071
148	LYS+	-	-0.568	-0.213

PCS of N60D CaM				
	Residue	**Tb**	**Tm**	**Dy**
4	LEU	-0.039	-	1.629
5	THR	1.074	-0.589	1.146
6	GLU	1.184	-0.609	1.095
7	GLU	1.092	-	0.966
8	GLN	-	-	0.771
9	ILE	1.750	-0.967	1.564
10	ALA	1.870	-1.062	1.66
11	GLU	1.426	-	-
12	PHE	1.749	-1.006	1.29
13	LYS+	2.751	-	1.87
14	GLU	1.886	-	1.008
15	ALA	2.352	-0.748	0.229
16	PHE	-	-	0.317
17	SER	2.031	-	-
18	LEU	0.009	-0.544	-
19	PHE	0.116	-0.054	-
21	LYS+	-1.338	-	-2.043
22	ASP	-0.685	-0.821	-
23	GLY	0.644	-1.491	-1.159
24	ASP	1.148	-2.372	-1.529
29	THR	-	-0.016	-0.027
30	LYS+	-4.582	-	-
31	GLU	-5.190	1.041	-
32	LEU	-7.254	-	-
33	GLY	-4.681	2.837	-
34	THR	-4.999	1.698	-2.129
35	VAL	-1.719	1.743	-
36	MET	-2.440	2.032	-2.63
37	ARG+	-2.037	-	-1.406
38	SER	-1.652	-	-1.455
39	LEU	-1.392	-	-
40	GLY	-1.182	0.798	-1.489
41	GLN	-1.344	-	-1.655
42	ASN	-2.073	-	-
44	THR	-0.949	1.217	-0.946
45	GLU	-	0.845	-
47	GLU	-1.364	-	-
48	LEU	-	2.231	-1.195
49	GLN	-	2.365	-
50	ASP	-2.379	2.435	-0.016
84	GLU	-	-	0.012

PCS of CLaNP-5-CaM			
	Residue	**Yb**	**Tm**
6	GLU	-0.029	-0.006
7	GLU	-0.033	-
10	ALA	-	0.019
13	LYS+	-0.003	-
15	ALA	-0.023	-
16	PHE	-	-0.011
18	LEU	-0.020	-0.065
19	PHE	-0.021	-
21	LYS+	-0.002	-0.040
22	ASP	-0.010	-0.027
23	GLY	-0.014	-
24	ASP	-0.017	-0.048
25	GLY	-	-0.044
26	THR	-0.012	-0.042
27	ILE	-0.022	-
28	THR	-	-0.058
29	THR	-0.018	-
30	LYS+	-0.003	-0.024
31	GLU	-0.014	-0.056
33	GLY	-0.029	-0.085
35	VAL	-	-0.068
38	SER	-0.011	-
44	THR	-0.025	-0.127
45	GLU	-0.007	-
46	ALA	-0.003	-0.020
48	LEU	-0.007	-0.076
49	GLN	-0.018	-
50	ASP	-0.023	-0.037
51	MET	-	-0.011
52	ILE	-0.022	-
53	ASN	-0.004	-0.011
54	GLU	-	0.001
55	VAL	-0.032	-
56	ASP	-	-0.037
57	ALA	-0.016	-0.023
58	ASP	-0.012	-0.010
59	GLY	0.010	0.014
60	ASN	-0.001	0.000
61	GLY	-0.027	-
62	THR	-	-0.011
64	ASP	-0.028	-0.085

85	ILE	-	0.124	-
86	ARG+	-	-	0.013
89	PHE	-0.095	0.070	0.002
90	ARG+	-	0.078	0.035
91	VAL	-0.134	0.007	0.064
93	ASP	-0.115	-	0.024
94	LYS+	-0.088	0.044	-
95	ASP	-0.061	0.029	0.043
96	GLY	-0.062	0.027	0.040
97	ASN	-0.056	0.037	0.027
98	GLY	-0.080	0.036	0.013
99	TYR	-0.064	0.047	0.017
100	ILE	-0.095	0.056	0.009
101	SER	-0.080	0.049	0.011
102	ALA	-0.066	0.048	0.002
103	ALA	-0.070	0.043	0.000
104	GLU	-0.035	0.048	0.007
105	LEU	-0.105	0.064	-0.004
106	ARG+	-	0.057	-0.026
107	HIS+	-	-	-0.021
109	MET	-0.115	-	-0.055
110	THR	-0.010	0.008	-
111	ASN	-0.002	0.000	0.000
112	LEU	-	-	-0.041
113	GLY	-0.011	0.013	-0.011
114	GLU	-	-	-0.010
115	LYS+	-0.002	0.017	0.001
116	LEU	-0.017	0.000	-0.068
117	THR	-0.102	0.074	-0.059
118	ASP	-0.078	0.058	-0.045
119	GLU	-0.065	0.061	-0.027
121	VAL	-	0.057	-0.040
122	ASP	-0.057	0.058	-0.017
123	GLU	0.017	0.051	-0.021
124	MET	-0.018	0.057	-0.023
127	GLU	-0.097	0.031	-0.029
128	ALA	-0.087	0.054	-0.008
129	ASP	-0.072	-	0.005
130	ILE	-0.090	0.042	-0.010
131	ASP	-0.078	0.036	-0.010
132	GLY	-0.048	0.050	-0.002
133	ASP	-0.043	-	-
134	GLY	-0.065	0.031	-0.005
135	GLN	-0.065	0.039	-0.003
136	VAL	-0.085	0.051	-0.001
137	ASN	-0.096	0.057	-0.002

65	PHE	-0.009	-
68	PHE	-0.014	-
70	THR	-0.029	-
86	ARG+	-	0.715
90	ARG+	-	0.904
91	VAL	-	1.511
93	ASP	0.523	-
94	LYS+	0.503	1.504
95	ASP	-0.140	-
96	GLY	-0.268	-
97	ASN	-0.071	-0.797
98	GLY	-0.066	-0.274
99	TYR	0.148	0.649
100	ILE	0.503	2.73
101	SER	0.702	3.137
102	ALA	0.561	2.26
103	ALA	0.645	-
104	GLU	1.118	3.867
105	LEU	1.467	-
113	GLY	1.565	-
115	LYS+	-	3.184
116	LEU	0.654	-
117	THR	0.337	2.02
118	ASP	0.250	-
121	VAL	0.546	-
122	ASP	-	1.812
123	GLU	0.364	-
124	MET	-	2.345
127	GLU	0.295	-
128	ALA	0.430	-
129	ASP	-	1.927
130	ILE	0.253	1.139
131	ASP	0.184	-
132	GLY	0.179	1.069
133	ASP	0.179	-
134	GLY	0.214	1.264
135	GLN	-	1.324
136	VAL	0.500	2.77
137	ASN	0.298	1.73
138	TYR	0.157	-
139	GLU	0.142	0.936
140	GLU	0.222	1.389
141	PHE	0.340	2.057
142	VAL	0.260	1.67
143	GLN	-	1.414
144	MET	0.351	1.833

138	TYR	-0.109	0.050	-0.010
139	GLU	-0.080	-	-
140	GLU	-0.118	0.069	0.001
141	PHE	-0.130	0.071	-0.003
142	VAL	-0.152	0.080	0.003
143	GLN	-0.143	0.081	0.000
144	MET	-0.037	-	-
145	MET	-0.163	0.080	-0.006
147	ALA	-0.151	0.107	0.021
148	LYS+	-0.147	0.087	0.017

145	MET	0.402	2.552

4
CONCLUSIONS AND PERSPECTIVE

The solution structures of S100A5 and S100A16 were solved in both apo and calcium bound states. The two proteins were found to exist as homodimer in both the apo and calcium bound forms.

S100A5 has the common characters of S100 proteins. It binds two calcium ions in each subunit. The structure analysis shows that it experiences large conformational changes upon the calcium ions binding: helix III rotates of about 50° with respect to helix IV and a cleft is formed at the hinge region between helix II and III, where the targets binding site is supposed to be. Also relaxation data suggest that in the hinge region there is more mobility than in the other parts of the proteins in both the apo and the calcium bound form.

S100A16 is a peculiar S100 member: it binds two calcium ions with lower binding affinity than the common S100 proteins, like S100A5, and consequently performs much smaller conformational changes upon calcium binding than S100A5. The relaxation data show that the first calcium binding loop and the beginning of the second helix experience more mobility than the other regions. This special dynamic property suggests that S100A16 could have different targets binding modes from other S100 proteins. More studies are needed to disclose its functional features.

S100A5 has a longer helix IV in the calcium bound state than in the apo form, analogously to S100A2, S100A4 and S100A6. On the contrary, in S100A16, helix IV has the same length in both the apo and the calcium bound forms because of a special sequence motif (Gly-Gly-Ile-Thr-Gly-Pro), which is adverse to the formation helix structures.

According to the available structures of S100 proteins, we measured their inter-helix angles and analyzed their conformational properties in both the apo and the calcium bound forms by using a principal component analysis. All the structural conformations are divided into two sub-groups, one for the close state and the other for the open state. Exceptions are S100A10 and calbindin $D_{9k.}$ S100A10 does not bind calcium and has an open state conformation in the metal free form; calbindin D_{9k} has open state conformation in both apo and calcium bound forms. While S100A5 is correctly positioned in the two groups according to the calcium bound or calcium-free state, S100A16 has a closed state conformation in both the apo and the calcium bound forms. Based on this structural information, energetics models for the calcium-triggered conformational changes of S100

proteins were made.

The structural details and the electrostatic potential surface obtained for S100A5 and S100A16 as well as their changes upon calcium binding are expected to reflect their different target specificity. Therefore, the knowledge of these features can provide information on the possible different function of the proteins.

From the NMR titration, the hMBF1 and CaM do not bind to each other *in vitro* both in the presence and in the absence of calcium(II). Further more, CaM- agarose study suggested that the binding of hMBF1 to the column is not dependent on the calcium-free state of CaM but depends on ionic interactions with the activated matrix, which can be prevented by the presence of salts. This result indicates that attention should be paid when investigating the interaction between proteins with column.

Synthetic tests and experimental studies performed on flexible two-domain proteins suggested that inter-domain motions can be better studied using paramagnetic restraints obtained with metal ions coordinated to both protein domains rather than in a single domain. In the case of CaM we have shown that the maximum occurrence of several conformations is quite reduced by the addition of pcs and rdc arising from the presence of a single metal ion rigidly attached to the C-terminal domain with respect to the values calculated using three metal ions placed in the N-terminal domain of the protein. As a result, the conformations likely experienced by the protein can be more accurately mapped.