University of Florence

International Doctorate in Structural Biology

Cycle XXIII (2007-2010)

Structural and Functional Characterisation of Calcium Sensing Proteins

PhD thesis of

Soumyasri Das Gupta

Tutor
Prof Claudio Luchinat

Coordinator
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.

Acknowledgements

First of all, I would like to express my sincere gratitude to Prof. Ivano Bertini, Dicrector, Magnetic Resonance Center (CERM) for giving me an opportunity to conduct my doctoral work in this esteemed institute. He has ensured my academic, professional, financial and moral well-being all through my tenure at CERM.

My sincere thanks to my supervisor, Prof Claudio Luchinat, for his suggestions, guidance and encouragement.

I take this opportunity to thank Dr. Marco Fragai for his constant support, encouragement and for extending a helping hand at all times of need through all these years.

I would like to convey my sincere thanks to my colleagues Yuan Jing, Xiaoyu Hu, Malini Nagulapalli, Dr. Giacomo Parigi, Valentina Borsi. I would like to express my warm and sincere thanks to my co-worker Dr. Antonella Nesi. I have truly learnt from her her concepts and ideas have helped me to improve my scientific perceptions and skills. I would like to convey my heartlfelt thanks to all my labmates at CERM. I really enjoyed working with them and will always cherish these years of my life.

Our technicians Leonardo Gonnelli and Dr. Marco Allegrozzi deserve my sincere thanks for their constant and selfless assistance in solving various problems.

Life in Florence would have been less pleasureable without my colleagues Rahul Jaiswal, Deepa Jaiswal, Dambrudhar Hembram, Vaishali Sharma, Ravikrishnan Elangovan, Malini Nagulapalli, Vasanth Kumar. My sincere thanks go to all of them for making me feel at home in this far off land.

I take this opportunity to thank the administrative and secretarial staff of CERM, Laura Norfini, Simona Fedi, Lisa Orlando, Milena Moazzi, Michele Natalini for extending their help in the complicated bureaucratic and secretarial work.

I am very grateful to Prof. Sunil Brahmachari, Vice-Chancellor, Mohonpur Agricultural University for being a source of inspiration to me.

I am deeply indebted to my parents and my family for their love, support and belief in me. I would like to thank my late mother with all my heart for whatever she has been to me. My special gratitude goes to my father-in-law for his constant encouragement.

Finally, I owe my most heartfelt and loving thanks to my husband Dr. Anusarka Bhaumik for being a friend, philosopher and guide to me. His support, encouragement and guidance have been of invaluable help in my scientific career.

Dedicated to my family and my husband….

Contents

List of Abbreviations

CaBP	Calcium binding protein
RAGE	Receptor for advanced glycation end products
CaM	Calmodulin
NMR	Nuclear magnetic resonance
PCR	Polymerase chain reaction
cDNA	Complementary DNA
IPTG	isopropyl-ß-galactosidase
PAGE	Polyacrilamide gel electrophoresis
CLaNP-5	Caged Lanthanide NMR Probe 5
PCS	Pseudo-contact shifts
RDC	Residual dipolar coupling
HSQC	Heteronuclear single quantum coherence

1. Introduction

1. Introduction

After the publication of genomes[1-5] of many organisms[6], including humans[3], the next collective challenge is to assign and understand the functions of the final genome products- proteins. The Structural Biology Roadmap is an effort to create a 'picture' gallery of the molecular shapes of proteins in the body. These three-dimensional insights are crucial for an understanding of basic life-processes, such as the reaction mechanism of a drug-converting enzyme, signal transduction from one protein to another, activation of a metabolic pathway by a gene effector, or the consequences of mutation on the function of an enzyme.

1.1 Calcium Signalling

Calcium (Ca^{2+}) is a ubiquitous intracellular signal responsible for controlling numerous cellular processes[7, 8]. Despite tremendous diversities in their expression, cellular activities in virtually all cell types are regulated by common intracellular signalling systems. Calcium is one important ubiquitous intracellular messenger, controlling a diverse range of cellular processes such as gene transcription, muscle contraction, cell proliferation[9-15]. At one level, its action is simple: cells at rest have a Ca^{2+} concentration of 100 nM[16] but are activated when this level rises to roughly 1000 nM[17].

An important challenge is to understand how this regulation of Ca^{2+} can regulate so many divergent cellular processes. This is possible due to the versatility of calcium signalling mechanism in terms of speed, amplitude and spatio-temporal patterning[18-19]. The versatility emerges from the use of an extensive molecular repertoire of signalling components which comprise a Ca^{2+}-signalling toolkit[20] that can be assembled in combinations to create signals with widely different spatial and temporal profiles. More variations are achieved through the interactions (crosstalk) that Ca^{2+} makes with other signalling pathways. This versatility is exploited to regulate various cellular processes.

The calcium signalling network can be divided into four functional units[21]:

- Signalling is triggered by a stimulus that generates various $Ca2^{+}$-mobilizing signals.
- The latter activates the ON mechanisms that feed Ca^{2+} into the cytoplasm.
- Ca^{2+} serves as a messenger to stimulate numerous Ca^{2}-sensitive processes.

- Finally, the OFF mechanisms composed of pumps and exchangers, remove Ca^{2+} from the cytoplasm to restore the resting state.

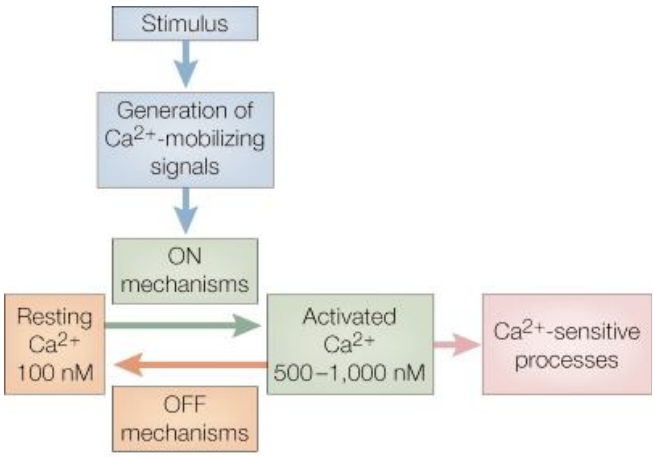


Figure 1: The four units of Ca^{2+} signalling network. Stimuli act by generating Ca^{2+} mobilising signals that act on various ON mechanisms to trigger an increase in intracellular concentration of Ca^{2+}. The increased level of Ca^{2+} stimulates various Ca^{2+}-sensitive processes to trigger many different cellular pathways. The response is terminated by OFF mechanisms that restore Ca^{2+} to its resting level (Ref21).

The functional relationship between these units is illustrated in the figure below which reveals that the signalling network is composed of many components (the Ca^{2+}-signalling toolkit). Because many of the molecular components of the toolkit have several isoforms with subtly different properties, each cell type can exploit this large repertoire to construct versatile Ca^{2+}-signalling networks[22].

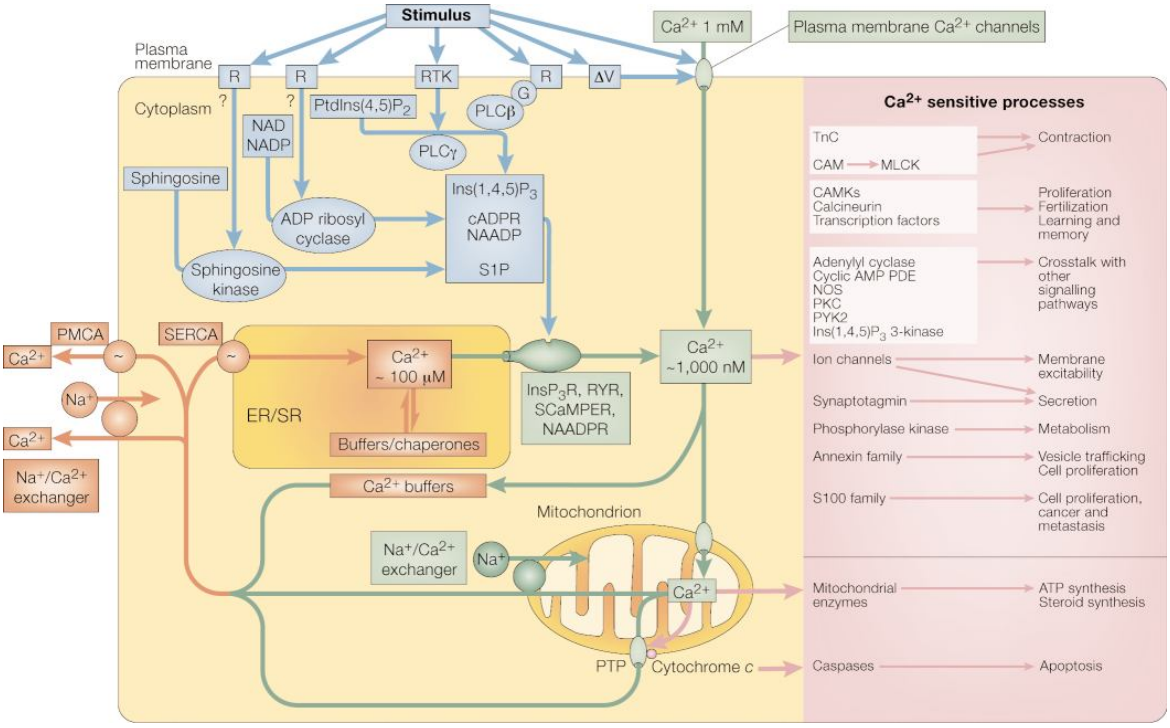


Figure 2: Elements of the Ca^{2+}-signalling toolkit : Cells have an extensive signalling toolkit that can be mixed and matched to create Ca^{2+} signals of widely different properties. (Ref. 21)

Each cell type expresses a unique set of components from the Ca^{2+}-signalling toolkit to create Ca^{2+} dependent signalling systems with different spatial and temporal properties[23]. Almost all Ca^{2+} signalling systems have one thing in common: they function by generating brief pulses of Ca^{2+} [24, 25]. These Ca^{2+} transients are created by variations of the basic ON/OFF reactions outlined in Fig.1.

Signal Ca^{2+} is derived either from internal stores or from external medium. In case of the latter, there are many different membrane channels defined by the way in which they are activated, like Voltage Operated Channels[26], Receptor Operated Channels[27], Store Operated Channels[28, 29]. The release of Ca^{2+} from the internal stores-mostly the endoplasmic reticulum (ER) or its muscle equivalent sarcoplasmic reticulum (SR)[30-32] is controlled by Ca^{2+} itself, or by a group of messengers such as innositol-l,4,5-triphosphate, cyclic ADP ribose, etc.

During the ON reaction, Ca^{2+} flows into the cell and interacts with different calcium binding proteins (CaBPs) to initiate cellular responses. CaBPs can be divided into two types on the basis of their functions: **Ca^{2+} buffers** and **Ca^{2+} sensors**[33]. The buffers which become loaded with Ca^{2+} during the ON reaction and unload during the off-reaction

function to fine-tune the spatial and temporal properties of Ca^{2+} signals[34]. Calbindin, calretinin, parvalbumin are examples of Ca^{2+} buffers[13]. Of the Ca^{2+} which enters the cytosol, only a very small proportion ends up being free, because most of it is rapidly bound to the buffers, and to a lesser extent to the sensors[35]. The later respond to an increase in Ca^{2+} by activating diverse processes. Typical examples of calcium sensors are troponin C, calmodulin, S100 proteins which have EF-hands that bind Ca^{2+} and undergo large conformational change to activate various downstream effectors[36-38].

The versatility of Ca^{2+} signalling is greatly enhanced by some of the Ca^{2+}-sensitive processes linking to other signalling pathways. The ability of Ca^{2+} to recruit the components of other signalling pathways (for example cyclic AMP and mitogen activated protein kinase pathways) [39, 40] is particularly evident in the control of gene transcription.
Once Ca^{2+} has carried out its signalling functions, it is rapidly removed from the cytoplasm by various pumps and exchangers. The plasma membrane Ca^{2+} ATPase pumps and Na^{+}/ Ca^{2+} exchangers extrude Ca^{2+} to the outside whereas the sarco-endoplasmic reticulum ATPase pumps return Ca^{2+} to the internal stores[41]. The mitochondrion is another important component of the OFF mechanism in that it sequesters Ca^{2+} rapidly during the development of Ca^{2+} signal and then releases it back slowly during the recovery phase.

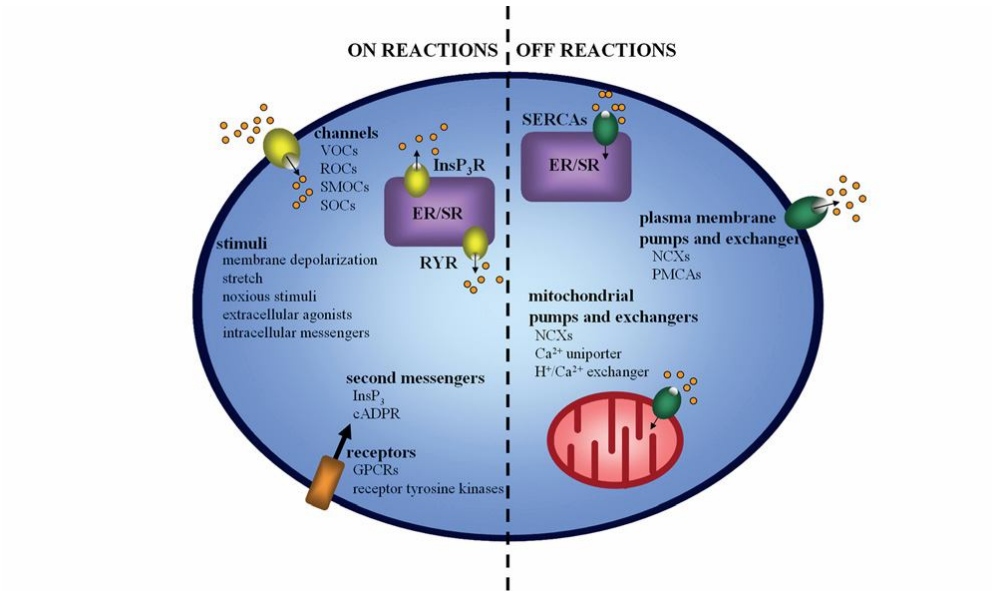


Figure 3: Ca^{2+} signalling in the cell

1.2 EF-hand Proteins

The EF-hand Ca^{2+} binding motif plays an essential role in eukaryotic cellular signalling, and the proteins containing this motif constitute a large and functionally diverse family, known as the EF-hand protein superfamily[42]. It is one of the most successful metal binding domains in the whole proteome, with more than seven thousand EF-hand proteins[43] annotated in the genome. Named after a Ca^{2+} binding motif first seen to be formed by helices E and F in the crystal structure of parvalbumin[44], the EF-hand is composed of a typical helix-loop-helix structural unit: two α helices bridged by a Ca^{2+} chelation loop.

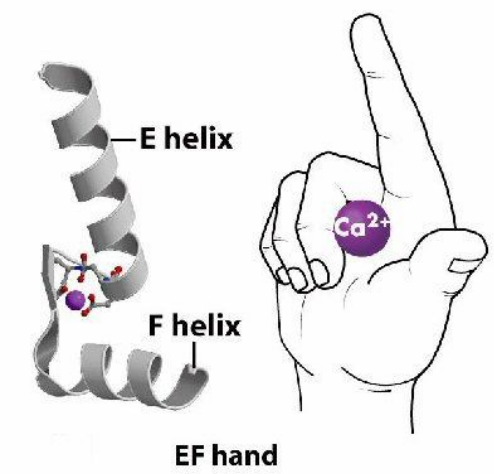


Figure 4: EF hand motif in calcium binding proteins

The identity of the ligands is semi-conserved in the most common (the 'canonical') EF-hand. However, several non-canonical EF-hands exist that bind Ca^{2+} by a different coordination mechanism[45]. EF-hands tend to occur in pairs which form a discrete domain so that most family members have two, four or six EF-hands[46]. The canonical EF-hand motif is characterised by a sequence of two helices of about 10 amino acids interrupted by a 12-residue loop that contains mainly polar and acidic residues where Ca^{2+} can be coordinated[47]. The conformational effects of Ca^{2+} binding are varied, function-dependent, and in some cases minimal, but can lead to the creation of a protein-target interaction site. The Ca^{2+} binding affinities of EF-hands vary substantially (K_d = 10^{-4}-10^{-9} M) and are amino-acid sequence dependent[48], especially with regard to the 12-residue consensus loop sequence that provides the ligands of the Ca^{2+} ion.

1.3 S100 Proteins

1.3.1 General Overview

The S100 proteins (soluble in 100 % ammonium sulphate at neutral pH) [49-50], with 25 members constitute the largest subgroup of EF-hand proteins. They are small (10-12kD), acidic proteins found exclusively in vertebrates[51]. These proteins are characterised by a high degree of conservation in amino acid sequence (25-65%) [52], 3-D structure[53] and genomic structure[54-55]. S100 proteins occur extra- and intra- cellularly exhibiting a cell and tissue-specific expression. They have no enzymatic activity of their own and exert their intracellular effects by interacting with and modulating the activities of other proteins, referred to as target proteins. Alteration of S100 functions have been implicated in a large number of diseases including cancer, Downe's syndrome, Alzheimer's disease, cardiomyopathy, psoriasis, cystic fibrosis, amyotrophic lateral sclerosis[56-61]. Thus S100 proteins may be important diagnostic markers as well as therapeutic targets for many diseases.

Most S100 genes are located in a cluster in human chromosome 1q21[62, 63], which is structurally conserved during evolution[64]. Within this chromosomal region, several rearrangements occur during tumour development[65]. This might be linked to a deregulation of S100 gene expression in various tumour types and is associated with metastasis and tumour development.

Another unique feature is that individual members of S100 proteins are localised in specific cellular compartments from which some of them are able to relocate upon Ca^{2+} activation[66-69], transducing the Ca^{2+} signal in a temporal and spatial manner, interacting with different targets specific for each S100 protein. Furthermore, some S100 proteins are even secreted from cells to exert cytokine- and chemokine-like extracellular activities. The individual members of S100 family seem to utilise distinct pathways (ER-Golgi route, tubulin or actin associated) for their secretion into the extracellular space[70].

1.3.2 Structural Overview

S100 proteins are found as non-covalent homodimers, heterodimers or oligomers, with the exception of S100G (calbindin) which is monomeric[71]. Each S100 monomer is characterised by the presence of two Ca^{2+} binding motifs of the EF-hand type interconnected by an intermediate region often referred to as the hinge region. In each

Ca^{2+} binding motif of the EF-hand type, a Ca^{2+} binding loop is flanked by two α-helices, resulting in a helix-loop-helix arrangement[72].

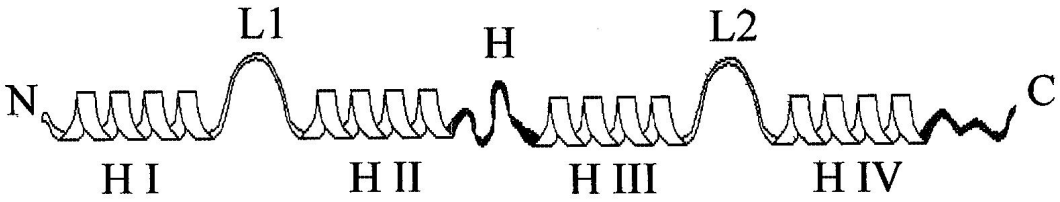


Figure 5: Schematic representation of the secondary structure of an S100 protein. Each Ca^{2+} binding loop (L1 and L2 in the N-and C-terminal half respectively) is flanked by α-helices (helices I and II and helices III and IV, for L1 and L2 respectively). A linker region, (hinge region, H) connects helix II-III. Helix IV is followed by a C-terminal extension. The hinge region and C-terminal extension display the least amount of homology among different members of the S100 family.

In case of S100 proteins, the first Ca^{2+} binding loop (EF1) is unconventional, in that it is longer (14 amino acids)[73] and rearranged where as the second Ca^{2+} binding loop is canonical (12 amino acids). Hence the two Ca^{2+} binding sites in a S100 protein bind Ca^{2+} with different affinities, EF1 or the pseudo EF hand having a lower affinity (K_D = 200-500 μM) than EF2 (K_D = 20-50 μM)[74]. Some S100 members also bind Zn^{2+} and Cu^{2+}. EF2 is followed by a C-terminal extension. The highest sequence identity among S100 members is found in the Ca^{2+} binding sites. The hinge region and the C-terminal extension exhibit the least amount of sequence identity suggesting that these two regions may have a role in the specification of biological activity of individual S100 proteins[75]. Upon Ca^{2+} binding almost all S100 proteins undergo a conformational change exposing a previously covered hydrophobic patch of residues[76]. There is a large change in the position of helix III upon Ca^{2+} binding. The interhelical angle between helix III and helix IV changes compared to the Ca^{2+} free structure, exposing the residues required for target recognition and binding[77].

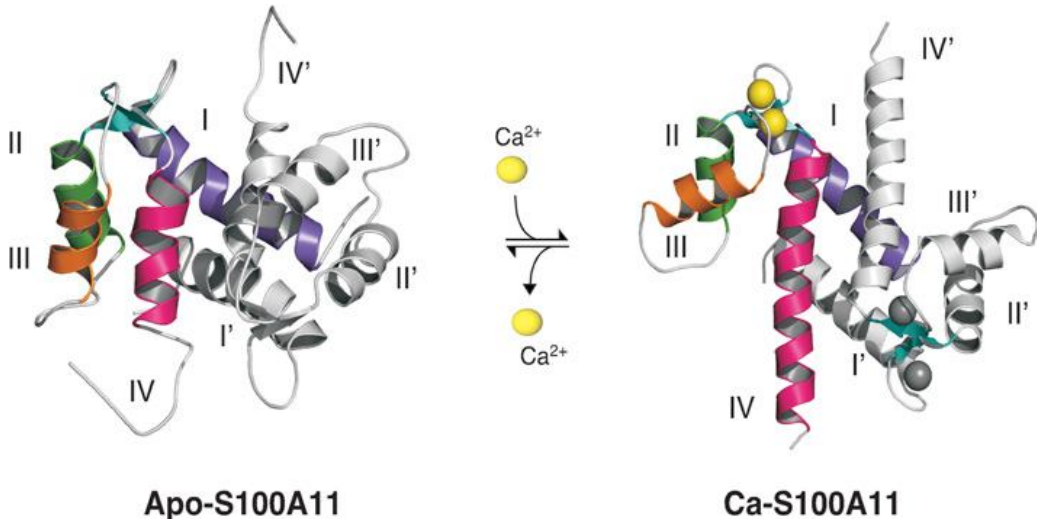


Figure 6: Calcium-dependent conformational changes in S100 proteins: The 3D structures of apo-S100A11 and calcium-S100A11 are shown to demonstrate the calcium-induced conformational change. In the symmetrical dimmer helices of monomer (I-IV) are highlighted in different colours, while the monomer (helices I'-IV') is coloured grey. As sensors, the S100 proteins experience a conformational change upon calcium binding (4 atoms/dimer). The rearrangement of helix III (orange) exposes previously buried residues that are essential to target recognition and further biological response. (Ref 77).

1.3.3 Functional Overview

S100 proteins are proposed to have intracellular and extracellular roles. In mature tissues, no S100 protein appears to be ubiquitous. Thus, a certain degree of cell specificity exists for any S100 member, implying that the relatively large no. of members is simply not due to redundancy. Members of this family are not exchangeable, or specific cell types need a specific S100 member. However, in some cases S100 proteins share their target proteins and, hence, regulate identical activities (possibly in different cell types). For eg., S100A1, S100A2, S100A4, S100B interact with the tumour suppressor p53 but in different ways[78-81]. In still other cases different S100 members take part in regulation of similar groups of activities. The intracellular roles of S100 proteins include regulation of protein phosphorylation[82-84], enzyme activity[85-88], Ca^{2+} homeostasis[89], cytoskeletal components[90-92] and transcription factors[92-93]. Some S100 members like S100A1, A2, A4, A12, S100B, S100P are secreted from cells upon stimulation (by a yet unknown mechanism), exerting cytokine and chemokine-like extracellular activities, via the

Receptor for Advanced Glycation End Products (RAGE). RAGE is a multiligand receptor of the immunoglobulin superfamily[94]. It has one V-type domain, two C-type domains, a transmembrane domain and a cytoplasmic tail. RAGE interacts with structurally different ligands probably through oligomerisation of the receptor on the cell surface[95]. Different S100 proteins appear to interact with distinct domains of the extracellular portion of RAGE to trigger various cellular effects[96]. RAGE is expressed at low levels in a wide range of differentiated adult cells[97]. It is highly expressed and associated with many inflammation-related pathological states such as chronic inflammation, diabetes, cancer, neurodegenerative disorders like Alzheimer's disease and multiple sclerosis[98, 99] along with S100 proteins. RAGE is a link between inflammatory pathways and pathways promoting tumourogenesis and metastasis. Therefore, characterising the role of S100-mediated RAGE signalling can be broadly applied to a variety of pathological conditions and incorporated into a wide array of treatment regimes for these conditions.

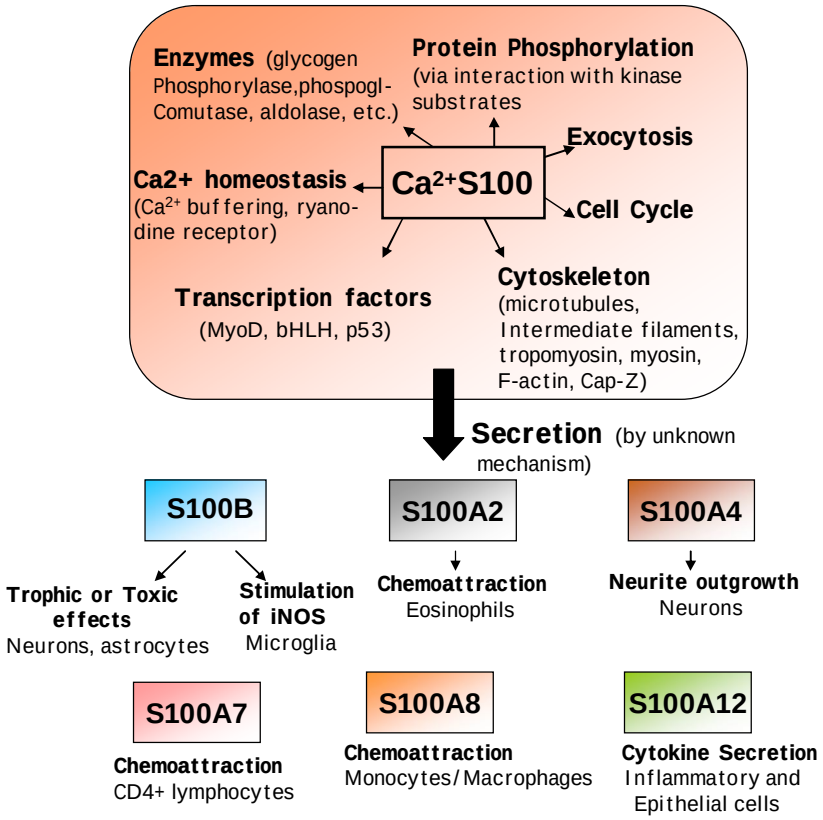


Figure 7: Schematic representation of extracellular and intracellular roles of S100 proteins

1.4 Receptor for Advanced Glycation End Products

RAGE is a member of the immunoglobulin superfamily of cell surface molecules[100, 101] and shares structural homology with other immunoglobulin like receptors[102, 103]. Although RAGE is not important to life[104], it plays important role in certain human pathologies including diabetes, Alzheimer's disease and cancer[105]. The mature 382 amino acid long RAGE is composed of an extracellular part (314 amino acids), a single transmembrane spanning helix (27 amino acids) and a short cytosolic domain (41 aa)[106]. The extracellular part of RAGE contains one variable like V-domain (residues 24-127) and two constant like C type domains frequently referred to as C1 (132-230) and C2 (239-320) domains. RAGE is known to form oligomers at the cell surface[107]. RAGE possesses two N-glycosylation sites, one adjacent to the V-domain and another within the V-domain[106, 108].

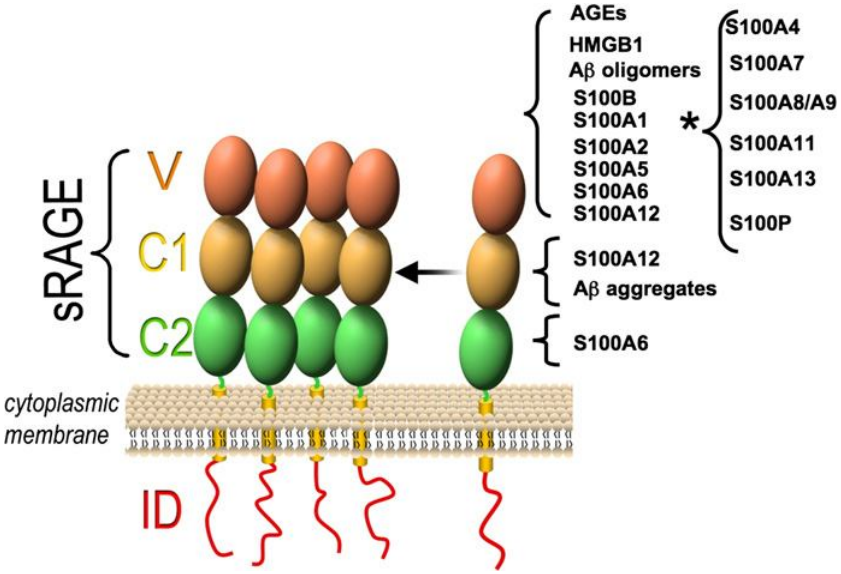


Figure 10: Schematic representation of RAGE. The extracellular part comprises V, C1 and C2 domains. A transmembrane helix connects the extracellular part with the short intracellular domain (ID). RAGE ligands interact with the extracellular domains as indicated[109].

RAGE ligands include structurally unrelated molecules like AGEs, amphoterin, amyloid ß peptide, immunoglobulin light chain amyloid fibrils, transthyretin, S100 proteins[110-115]. RAGE is highly expressed during development but its expression level decreases in adult tissue except in lung tissue where it is found in high level[116]. RAGE expression is also

augmented by increased levels of ligands in pathological states. RAGE signalling is complex and depends on the cell type, the type and concentration of the ligand. Understanding the transcriptional regulation of RAGE is important to understand RAGE signalling. Based on mRNA splicing, 20 isoforms of RAGE have been identified so far in various tissues and cells. Many of the S100 proteins (A1, A2, A4, A5, A6, A8/A9, A12, S100B, S100P) interact with RAGE and trigger RAGE dependent signalling pathways. Activation of multiple intracellular signalling molecules including the transcription factor, NF-kB, MAP kinases and adhesion molecules are noted following activation of RAGE. S100 proteins use different mechanisms for binding to RAGE. For example, S100B is a homotetramer in the high Ca^{2+} extracellular environment and on binding to RAGE mediates dimerisation of the receptor[117]. S100A12, on the other hand, occurs as hexamer and causes RAGE tetramerisation[118]. RAGE signalling is implicated in inflammation and tumour development.

1.5 Calmodulin

Calmodulin represents the prototypical intracellular Ca^{2+} sensor containing four Ca^{2+} binding sites in the loops of four canonical EF-hand motifs[119]. It is highly conserved and widely distributed in all members of the animal and plant kingdoms, fungi and protozoa with 100% amino acid sequence identity among all vertebrates that synchronise cellular responses to cell activation, resulting from an elevation of Ca^{2+} [120-122].

CaM is a 149 amino acid protein consisting of two globular domains, each containing two EF-hand motifs connected by a central helix.

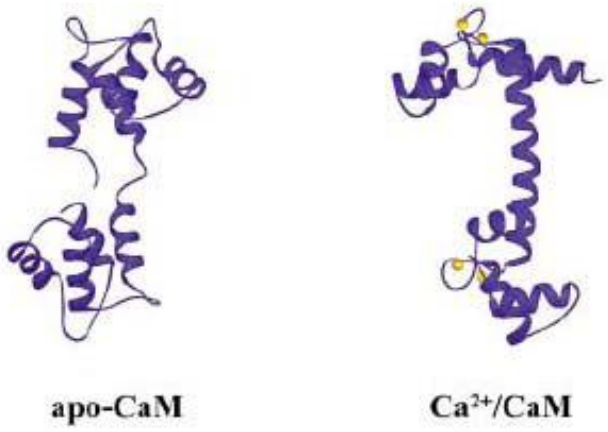


Figure 11: Structure of CaM in the apo and Ca4 forms. Upon binding Ca^{2+}, the central helix of CaM connecting the N-and C-terminal domains bends round leading to a

conformational change in the overall structure of CaM, and to the exposure of hydrophobic residues interacting with target proteins.

The pairing of EF-hands enables cooperativity in their binding of Ca^{2+} ions[123]. The two domains share high overall sequence homology (75%), as well as structural similarity in the presence or absence of Ca^{2+} ions. However, two Ca^{2+} ions bind with a 10-fold lower affinity (Kd = 10^{-5} M) to the N-domain than to the C-domain (Kd = 10^{-6} M). This allows CaM to sense transient Ca^{2+} variations in the cytoplasm over a relatively wide concentration range. Upon Ca^{2+} binding, the linker between the two domains bends round and CaM assumes a more globular shape ready to wrap around a substrate recognition site. At the same time, the two domains undergo conformational changes exposing hydrophobic patches that favour target protein interactions[124].

CaM can bind to its targets in different ways; in the extended mode interaction its domains bind with different regions of the target[125]. Complexes of CaM with proteins from the family of bHLH transcription factors reveal another binding mode that leads to CaM-induced dimerisation of the target[126].

The flexibility of CaM structure and the different binding modes are the key features that enable CaM to interact with more than hundred different targets, involved in numerous cellular processes like cell division, differentiation, transcription, ion transport by membrane channels, membrane fusion, and muscle contraction[127-129].

1.6 Aims and Overview of the Research

Conversion of the biological information carried by an external or internal stimulus into a biological function is a result of the concerted action of various toolkits present inside the cells. Proteins are often represented as the major components of such toolkits and some proteins have dual or multiple roles as communicators and act as hubs between these toolkits. Calcium signalling is one of the major cellular events. S100 proteins constitute the largest family of calcium-sensing proteins of the EF-hand type and control diverse cellular processes through interaction with an array of target proteins. Despite their similar amino-acid sequences and 3-D structures, individual family members exhibit diverse biophysical/biochemical properties. However, little is known about the molecular mechanisms of their action. Calmodulin, the most versatile protein in vertebrates is able to modulate a number of biochemical pathways and acts as a hub for the toolkits that participate in calcium mediated signalling events. During my PhD I have focussed on the cloning, expression, purification and characterisation (structural and functional) of the calcium sensing proteins of S100 family and calmodulin.

The physiological function of a protein is strictly determined by its function. Solution NMR is a powerful tool for structure determination and mobility studies of proteins. One aspect of my research involves determination of solution structure of the protein S100A5 and its dynamics studies in apo- and $Ca2^{+}$-loaded forms.

S100 proteins have been implicated in the Ca^{2+} dependent regulation of various intracellular activities like protein phosphorylation, cell proliferation, differentiation, motility, etc. Some members are released or secreted into the extracellular space, exerting cytokine-like activities through interaction with the cell surface receptor, RAGE (receptor for advanced glycation end products). S100 mediated RAGE signalling is implicated in diabetes, chronic inflammation, tumour progression. Another line of my research was based on interaction studies of RAGE with several extracellular S100 proteins- S100A1, A2, A4, A5, A12, S100P. Interactions of S100 proteins with different domains of soluble RAGE were studied by solution NMR. For structural characterisation of the complex, a homology model was obtained for S100P with the VC1 domain of RAGE and the interaction surface mapped. The calcium binding sites of S100 proteins were mutated to attach a paramagnetic probe to them for structural investigation of the complex. For functional characterisation of S100-RAGE complexes, two human cancer cell lines Panc1 and SW480 which over-express RAGE were treated with different concentrations of

recombinant extracellular S100 proteins. The dose-dependent effect of S100 treatment on cell proliferation was studied. Our findings suggest that S100 proteins stimulate proliferation of these cancerous cell lines by 20-60%, explaining the role of these proteins in tumour progression. Competition experiments were performed where Panc-1 cell line was treated with S100 proteins in the presence of different domains of soluble RAGE. The objective was to find out if treating cells with a S100 protein in presence of the RAGE domain with which it interacts prevented it from stimulating cell proliferation.

A third line of my research aims at determining the conformational space sampled by the two domain flexible CaBP, CaM based on the use of lanthanide ions in both the N-and C-terminal domains and collection of pseudocontact shifts (PCS) and residual dipolar coupling (RDC) induced by them. The N60D CaM mutant was used to selectively substitute the Ca^{2+} with paramagnetic ions at the N-domain. A paramagnetic probe was introduced in the C-domain of a CaM mutant (with two cysteines) through a lanthanide binding tag. The pcs and rdc values calculated for the moving domain with respect to the fixed (lanthanide-bearing) domain were used to analyse the interdomain conformational variability. The CaM conformations with largest MO (maximum occurrence) were calculated using these restraints. MO strategy is based on determining the maximum weight that a given conformation can have in any conformational ensemble in agreement with all available experimental data. The MO of each conformation decreases towards actual probability when the number of experimental restraints is increased. This project aimed to analyse how MO values are affected by restraints provided by paramagnetic ions located in both domains of CaM. The uniqueness of this study lies not only in the structural information about a protein which samples multiple conformations but is also in terms of the approach developed to achieve the results. The same approach should, in principle, be applicable to other multidomain proteins, as well as to multiple interaction modes between two macromolecular partners. This type of information is particularly desirable for a biologically significant protein like calmodulin, which is present in all eukaryotes and interacts with hundreds of targets to regulate multiple cellular pathways.

1.7 References

1. Downes, A.M. & Richardson, B.J. Relationship between genomic base contents and distribution of mass in coded proteins. *J. Mol. Evol.* **55,** 476-490 (2002).
2. Sherstnev, V.V. Neurospecific regulatory genome proteins in brain cells and systemic processes of memory. *Vestn. Ross. Akad. Med. Nauk.* 16-19 (1994).
3. Sorenson, D.K. Human genome protein function database. *Proc. Annu. Symp. Comput. Appl. Med. Care.* 434-438 (1991).
4. Stuart, G. W., Moffet, K. & Baker, S. Integrated gene and species phylogenies from unaligned whole genome protein sequences. *Bioinformatics* **18,** 100-108 (2002).
5. Yao, T. World trends of systems biology related to genome network project. *Tanpakushitsu Kakusan Koso* **49,** 2993-3000 (2004).
6. Yazaki, K. Morphology of genome and genome-protein complexes of viruses. *Tanpakushitsu Kakusan Koso* **37,** 2484-2492 (1992).
7. Berridge, M. J., Lipp, P. & Bootman M. D. The versatility and universality of calcium signalling. *Nat. Rev. Mol. Cell. Biol.* **1,** 11-21 (2000).
8. Carafolli E., Santella, L., Brance D. & Brisi, & M. Generation, control and processing of cellular calcium signals. *Crit. Rev. Biochem. Mol. Biol.* **36,** 107-260 (2001).
9. Hofer A. M. Another dimension to calcium signalling: a look at extracellular calcium. *J. Cell Sci.* **118,** 855-862 (2005).
10. Haug-Collet, K., Pearson, B., Webel, R., Szerencsei, R.T., Winkfein, R. J. & Schnetkamp, P. P. M. Cloning and characterisation of a potassium dependent sodium/calcium exchanger in *Drosophila*. *J. Cell. Biol.* **147,** 659-669 (1999).
11. Yuzo F., Lundmo, P., Short A.D., Gill, D.L. & Isaacs, J.T. The role of calcium, pH and cell proliferation in the programmed (apoptotic) death of androgen independent prostatic cancer cells induced by thapsigargin. *Cancer Res.* **54,** 6167-6175 (1994).
12. Ikura, M., Osawa, M. & Ames, J.B. The role of calcium binding proteins in the control of transcription: structure to function. *Bioessays* **24,** 625-636 (2002).
13. Enslen, H. & Soderling, T.R. Roles of calmodulin dependent protein-kinases and phosphatise in calcium dependent transcription of immediate early genes. *J. Biol. Chem.* **269,** 20872-20877 (1994).
14. Szent-Gyorgyi, A.G. Calcium regulation of muscle contraction. *Biophys J.* **15(7),** 707-723 (1975).

15. Berchtold, M.W., Brinkmeier, H. & Muentener, M. Calcium ion in skeletal muscle: its crucial role for muscle function, plasticity and disease. *Physiol. Rev.* **80**, 1215-1265 (2000).
16. Williams, R.J. Calcium binding proteins in normal and transformed cells. *Cell Calcium* **20,** 87-93 (1996).
17. Hofer, A.M. Another dimension to calcium signalling: a look at extracellular calcium. *J. Cell. Sci.* **118,** 855-862 (2005).
18. Dargan, S.L., Schwaller, B. & Parker, I. Spatiotemporal patterning of IP_3-mediated Ca^{2+} signals in *Xenopus* oocytes by Ca^{2+} binding proteins. *J. Physiol.* **556,** 447-461 (2004).
19. Bi, C.M., Dai, G., Chen, Y., Yu, W.C., Xhang, X.R. & Li, C.J. Spatio temporal patterning of calmodulin and Ca^{2+} is related to resumption of meiosis in mouse oocytes. *Cell Biol. Int.* **28(4),** 317-322.
20. Bootman, M,D., Collin T.J., Peppiatt, C.M., Prothero, L.S., MacKenzie, L., Smet, P.D., Travers, M., Tovey, S.C., Seo, J.T., Berridge, M.J., Ciccolini, F. & Lipp, P. Calcium signalling: an overview. *Semin. Cell. Dev. Biol.* **12,** 3-10 (2001).
21. Berridge, M.J., Lipp, P. & Bootman, M.D. The versatility and universality of calcium signalling. *Nat. Rev. Mol. Cell Biol.* **1,** 11-21 (2000).
22. Berridge M.J., Bootman, M.D. & Roderick, H.L. Calcium signalling: dynamics, homeostasis and remodelling. *Nat Rev Mol Cell Biol.* **4,** 517-529 (2003).
23. Van Ginkel, P.R., Gee, R.L., Walker, T.M., Hu, D.N., Heizmann, C.W. & Polans, A.S. The identification and differential expression of calcium binding proteins associated with ocular melanoma. *Biochim. Biophys. Acta* **1448,** 290-298 (1998).
24. Martin, B.J., Valdivia, H.H., Buenger, R., Lasley, R.D. & Mentzer, R.M. Pyruvate augments calcium transients and cell shortening in rat ventricular myocytes. *Am. J. Physiol. Heart Circ. Physiol.* **274,** 8-17 (1998).
25. Ciapa, B., Pesando, D., Wilding, M. & Whitaker, M. Cell cycle calcium transients driven by cyclic changes in innositol triphosphate levels. *Nature* **368,** 875-878 (1994).
26. Dildy-Mayfield, J.E., Machu, T. & Leslie, S.W. Ethanol and voltage-or receptor-mediated increases in cytosolic Ca^{2+} in brain cells. *Alcohol* **9,** 63-69 (1992).
27. Putney, J.W. A model for receptor-regulated calcium entry. *Cell Calcium* **7,** 1-12 (1986).
28. Fukushima, M.M., Tomita, T., Bird, G. S. & Putney, J.W. Store-operated Ca^{2+} entry in NIH-3T3 cells. *J. Med. Invest.* **56,** 381-382 (2009).

29. Venkatachalam, K., van Rossum, D.B., Patterson, R.L., Ma, H.T. & Gill, D.L. The cellular and molecular basis of store-operated calcium entry. *Nat. Cell Biol.* **4,** E263-E272 (2002).
30. Roderick, H.L. & Bootman, M.D. Redox calcium from the ER. *Cell* **120,** 4-5 (2005).
31. Balakier, H., Dziak, E., Sojecki, A., Librach, C., Michalak, M. & Opas, M. Calcium binding proteins and calcium release channels in human maturing oocytes, prenuclear zygotes and early preimplantation embryos. *Hum. Reprod.* **17(11),** 2938-2947 (2002).
32. Sordas, G. & Hajnoczky. SR/ER mitochondrial local communication. *Biochim. Biophys. Acta* **178 (11),** 1352-1362 (2009).
33. Braunewell, K.H. & Gundelfinger, E.D. Intracellular neuronal calcium sensor proteins: a family of EF-hand calcium binding proteins in search of a function. *Cell Tissue Res.* **295,** 1-12 (2009).
34. Berridge, M.J., Bootman, M.D. & Roderick, H.L. Calcium signalling: dynamics, homeostasis and remodelling. *Nat. Rev.* **4,** 517-529 (2003).
35. Zeller, S., Rudiger, S., Engel, H., Sneyd, J., Warnecke, G., Parker, I. & Falcke, M. Modelling the modulation by buffers of Ca2+ release through clusters of IP3 receptors. *Biophysics J.* **97(4),** 992-1002 (2009).
36. McCue, H.V., Haynes, L.P. & Burgoyne, R.D. The diversity of calcium sensor proteins in the regulation of neuronal function. *Cold Spring Harb. Perspect. Biol.* a004085 (2010).
37. Burgoyne, R.D. & Weiss, J.L. The neuronal calcium sensor family of Ca^{2+} binding proteins. *Biochem. J.* **353,** 1-12 (2001).
38. Schaub, M.C. & Heizmann, C.W. Calcium, troponin, calmodulin, S100 proteins: From myocardial basics to new therapeutic strategies. *Biochem. Biophys. Res. Comm.* **369,** 247-264 (2007).
39. Egea, J., Espinet, C. & Comella, J.X. Calcium influx activates extracellular-regulated kinase/mitogen activated protein kinase pathway through a calmodulin sensitive mechanism in PC_{12} cells. *J. Biol. Chem.* **274,** 75-85 (1999).
40. Gevrey, J.C., Bussat, M.C., Gaillard, E.N., Chayvialle, J.A. & Abello, J. Co-requirement of cyclic AMP and calcium-dependent protein kinases for transcriptional activation of cholecystokinin gene by protein hydrolysates. *J. Biol. Chem.* **227,** 22407-22413 (2002).
41. Edge, L. Calcium Signaling. *Cell* **131,** 1407-1058 (2007).

42. Yap. K.L., Ames, J.B., Swindells, M.B. & Ikura, M. Diversity of conformational states and changes within the EF-hand protein superfamily. *Proteins* **37(3),** 499-507 (1999).
43. Kawasaki, H., Nakayama, S. & Kretsinger, R.H. Classification and evolution of EF-hand proteins. *Biometals* **11,** 277-295 (1998).
44. Declercq, J.P., Evrard, C., Lamzin, V. & Parello, J. Crystal structure of the EF-hand parvalbumin at atomic resolution (0.91 Å) and at low temperature (100K). Evidence for conformational multistates within the hydrophobic core. *Protein Science* **8,** 2194-2204 (1999).
45. Gifford, J.L., Walsh, M.P. & Vogel, H.J. Structures and metal ion binding properties of the Ca^{2+} binding helix-loop-helix EF-hand motifs. *Biochem. J.* **405,** 199-221 (2007).
46. Nelson, M.R. & Chazin, W.J. Structures of EF-hand Ca^{2+} binding proteins: Diversity in the organisation, packing and response to Ca^{2+} binding. *Biometals* **11,** 297-318 (1998).
47. Capozzi, F., Casadei, F. & Luchinat, C. EF-hand protein dynamics and evolution of calcium signal transduction: an NMR view. *J. Biol. Inorg. Chem.* **11,** 949-962 (2006).
48. Ikura, M. Calcium binding and conformational response in EF-hand proteins. *Trends Biochem. Sci.* **21,** 14-17 (1996).
49. Moore, B.W. A soluble protein characteristic of the nervous system. *Biochem. Biophys. Res. Commun.***19,** 739-744 (1965).
50. Zimmer, D.B., Cornwall E.H., Landar, A. & Song, W. The S100 protein family: history, function and expression. *Brain. Res. Bull.* **37,** 417-429 (1995).
51. Donato, R. Intracellular and extracellular roles of S100 proteins. *Microsc. Res. Technol.* **60,** 540-551 (2003).
52. Schaefer, B.W. & Heizmann, C.W. The multifunctional S100 protein family. *Trends. Biochem. Sci.* **21,** 134-140 (1996).
53. Rustandi, R.R., Baldisseri, D.M. & Weber, D.J. Structure of the negative regulatory domain of p53 bound to S100B. *Nat. Struct. Biol.* **7,** 570-574, 2000.
54. Heizmann, C.W. The multifunctional S100 protein family. *Methods Mol. Biol.* **172,** 69-80 (2002).
55. Zimmer, D.B. Chessher J.C. & Sorg, W. Nucleotide homologies in genes encoding members of the S100 protein family. *Biochim. Biophys. Acta.* **13113,** 229-238 (1996).
56. Van Eldik, L.J. & Griffin, W.S.T. S100B expression in Alzheimer's disease: relation to neuropathology in brain regions. *Biochim. Biophys. Acta* **1223,** 398-403 (1994).

57. Schaefer, B.W. & Heizmann, C.W. The S100 family of EF-hand calcium binding proteins: functions and pathology. *Trends Biochem. Sci.* **21,** 134-140 (1996).
58. Heizmann, C.W. & Cox, J.A. New perspectives on S100 proteins: a multifunctional Ca^{2+}, Zn^{2+} and Cu^{2+} binding protein family. *Bio Metals* **11,** 383-397 (1998).
59. Kerkhoff, C., Klemp, M. & Sorg, C. Novel insights into structure and function of MRP8 (S100A8) and MRP14 (S100A9). *Biochim. Biophys. Acta* **1448,** 200-211 (1998).
60. Barraclough, R. Calcium binding protein S100A4 in health and disease. *Biochim. Biophys. Acta* **1448,** 190-199 (1998).
61. Donato, R. S100: A multigenic family of calcium modulated proteins of the EF-hand type with intracellular and extracellular functional roles. *Int. J. Biochem. Cell Biol.* **33,** 637-668 (2001).
62. Schaefer, B.W., Wicki, R., Engelkamp, D., Mattei, M.G. & Heizmann, C.W. Isolation of a YAC clone covering a cluster of nine S100 genes on human chromosome 1q21:rationale for a new nomenclature of the S100 protein family. *Genomics.* **25,** 638-643 (1995).
63. Marenholz, I., Zina, M., Fischer, D.F., Backendorf, C., Zeigler, A. & Mischke, D. Identification of human epidermal differentiation complex (EDC)-encoded genes by subtractive hybridisation of entire YACs to a gridded keratinocyte cDNA library. *Genome Res.* **11,** 341-355 (2001).
64. Ridinger, K., Ilg, E.C., Niggli, F.K., Heizmann, C.W. & Schaefer, B.W. Clustered organisation of S100 genes in human and mouse. *Biochim. Biophys. Acta* **1448,** 254-263 (1998).
65. Schutte, B.C., Carpten, J.D., Forus, A., Gregory, S.G., Horii, A. & White P.S. Report of the sixth international workshop on human chromosome 1 mapping 2000. *Cytogenet. Cell Genet.* **92,** 23-48 (2001).
66. Mandinova, A., Atar, D., Schaefer, B.W., Spiess, M., Aebi, U. & Heizmann, C.W. Distinct subcellular localisation of the calcium binding S100 proteins in human smooth muscle cells and their relocation in response to rises in intracellular calcium. *J. Cell Sci.* **111,** 2043-2054 (1998).
67. Dayey, G.E., Mumann, P & Heizmann, C.W. Intracellular Ca^{2+} and Zn^{2+} levels regulate the alternative cell-density dependent secretion of S100B in human glioblastoma cells. *J Biol. Chem.* **276,** 30819-30826 (2001).

68. Brett, W., Mandinova, A., Remmpis, A., Sauder, U., Rueter, F., Heizmann, C.W., Aebi, U. & Zerkowski, H.R. Translocation of S100A1 calcium binding protein during heart surgery. *Biochim. Biophys. Res. Commun.* **284,** 698-703 (2001).
69. Mueller, A., Bachi, T., Hochli, M., Schaefer, B.W. & Heizmann, C.W. Subcellular distribution of S100 proteins in tumour cells and their relocation in response to calcium activation. *Histochem. Cell. Biol.* **111,** 453-459 (1999).
70. Hsieh, H.L., Schaefer, B.W., Cox, J.A. & Heizmann, C.W. S100A13 and S100A6 exhibit distinct translocation pathways in endothelial cells. *J. Cell. Sci.* (2002).
71. Potts, B.C., Carlstrom, G., Okazaki, K., Hidaka, H. & Chazin, W.J. 1H NMR assignments of apo calcyclin and comparative structural analysis with calbindin D9k and S100B. *Prot. Sci.* **5(11),** 2162-2174 (1996).
72. Giford, J.L., Walsh, M.P. & Vogel, H.J. Structures and metal-ion-binding properties of the Ca^{2+} binding helix-loop-helix EF-hand motifs. *Biochem J.* **405,** 199-221 (2007).
73. Zimmer, D.B. & Weber, D.J. The calcium dependent interaction of S100B with its protein targets. *Cardiovasc. Psych. Neurol.* (2010).
74. Ishikawa, K., Nakagawa, A., Tanaka, I., Suzuki, M. & Nishihira, J. The structure of human MRP8, a member of the S100calcium-binding protein family, by MAD phasing at 1.9 Å resolution. *Acta Cryst.* **56,** 559-566 (2000).
75. Raftery, M.J., Harrison, C.A., Alewood, P., Jones, A. & Geczy, C.L. Isolation of the murine S100 protein MRP14 (14 kDa migration-inhibitory factor-related protein) from activated spleen cells: characterisation of post translational modifications and zinc binding. *Biochem. J.* **316,** 285-293, 1996.
76. Moroz, O.V., Burkitt, W., Wittkowski, H., He, W., Ianoul, A., Novitskaya, V., Xie, J., Polyakova, O., Lendev, I.K., Shektman, A., Derrick, P.J., Bjoerk, P., Foell, D. & Bronstein, I.B. Both Ca^{2+} and Zn^{2+} are essential for S100A12 protein oligomerisation and function. *BMC Biochemistry* **10,** 1471-2091 (2009).
77. Kisiel, L.S., Dempsey, A.C.R., Shaw, G.S. Calcium-dependent and independent interactions of the S100 protein family. *Biochem. J.* **396,** 201-214 (2006).
78. Mueller, A., Schaefer, B.W., Ferrari, S., Weibel, M., Makek, M., Hochli, M. & Heizmann, C.W. The calcium binding protein S100A2 interacts with p53 and modulates its transcriptional activity. *J. Biol. Chem.* **280,** 29186-29193 (2005).
79. Fernandez, M.R., Veprintsy, D.B. & Fersht, A.R. Proteins of the S100 family regulate the oligomerisation of p53 tumour suppressor. *Proc. Natl. Acad. Sci.* U.S.A. **102,** 4735-4740 (2005).

80. Grigorian, M., Andressen, S., Tulchinsky, E., Kriajevska, M., Carlberg, C., Kruse, C., Cohn, M. Ambartsumian, N., Christensen, A., Selivanova, G. & Lukanidin, E. Tumour suppressor p53 protein is a new target for the metastasis associated mts1/S100A4 protein. *J. Biol. Chem.* **276,** 22699-22708 (2001).
81. Chen, H., Fernig, D.G. Rudland, P.S., Sparks, A., Wilkinson, M.C. & Barraclough, R. Binding to intracellular targets of the metastasis-inducing protein, S100A4 (p9Ka). *Biochem. Biophys. Res. Commun.* **286,** 1212-1217 (2001).
82. Baudier, J., Mochly-Rosen, D., Newton, A., Lee, S.H., Koshland, D.E. & Cole, R.D. Comparison of S100B with calmodulin: interactions with mellitin and microtubule associated tau proteins by protein kinase C. *Biochemistry* **26,** 2886-2893 (1987).
83. Baudier, J. & Cole, R.D. Interactions between microtubule-associated tau proteins and S100B regulate tau phosphorylation by the Ca^{2+}/calmodulin-dependent protein kinase II. *J. Biol. Chem.* **263,** 5876-5883 (1988).
84. Sheu, F.S., Azmitia, E.C., Marshak, D.R., Parker, P.K. & Routtenberg, A. Glial-derived S100B protein selectively inhibits recombinant protein kinase C (PKC) phosphorylation of neuron-specific protein F1/GAP43. *Mol. Brain. Res.* **21,** 62-66 (1994).
85. Zimmer, D.B. & Van Eldik, L.J. Identification of a molecular target for the calcium modulated protein S100: fructose-1, 6-bisphosphate aldolase. *J. Biol. Chem.* **261,** 11424-11428 (1986).
86. Landar, A., Caddel, G., Chessher, J. & Zimmer, D.B. Identification of an S100A1/S100B target protein: phosphoglucomutase. *Cell Calcium* **20,** 279-285 (1996).
87. Millward, T.A., Heizmann, C.W. Schaefer, B.W. & Hemmings, B.A. Calcium regulation of Ndr protein kinase mediated by S100 calcium binding proteins. *EMBO J.* **17,** 5913-5922 (1998).
88. Margulis, A., Pazdnyakov, N. & Sitaramayya, A. Activation of bovine photo receptor guanylate cyclase by S100 proteins. *Biochem. Biophys. Res. Commun.* **218,** 243-247 (1996).
89. Zimmer, D.B. & Landar, A. Analysis of S100A1 expression during skeletal muscle and neuronal cell differentiation. *J. Neurochem.* **64,** 2727-2736 (1995).
90. Sorci, G., Agneletti, A.L., Bianchi, R. & Donato, R. Association of S100B with intermediate filaments and micro tubules in glial cells. *Biochim. Biophys. Acta* **1448,** 277-289 (1998).

91. Donato, R. Calcium-sensitivity of brain microtubule proteins in the presence of S100 proteins. *Cell Calcium* **6,** 343-361 (1985).
92. Donato, R. Calcium-independent, pH-regulated effects of S100 proteins on assembly-disassembly of brain microtubule protein in vitro. *J. Biol. Chem.* **263,** 106-110 (1988).
93. Baudier, J., Bergeret, E., Bertacchi, N., Weintraub, H., Gagnon, J. & Garin, J. Interaction of bHLH transcription factors with calcium-binding calmodulin and S100a proteins. *Biochemistry* **34,** 7834-7846 (1995).
94. Onions, J., Hemann, S. & Grundstroem, T. Basic helix-loop-helix protein sequences determining differential inhibition by calmodulin and S100 proteins. *J. Biol. Chem.* **272,** 23930-23937 (1997).
95. Dattilo, B.M., Fritz, G., Leclerc, E., Vander Cooi, C.W., Heizmann, C.W. & Chazin, W.J. The extracellular region of the receptor of advanced glycation end products is comprised of two independent structural units. *Biochemistry* **46(23),** 6957-6970 (2007).
96. Leclerc, E., Fritz, G., Vetter, S.W. & Heizmann, C.W. Binding of S100 proteins to RAGE: an update. *Biochim. Biophys. Acta* **1793,** 993-1007 (2009).
97. Xie, J., Burz, D.S., He, W., Bronstein, I.B., Lednev, I & Shektman, A. Hexameric calgranulin C (S100A12) binds to the receptor for advanced glycation end products (RAGE) using symmetric hydrophobic target binding patches. *J. Biol. Chem.* **282,** 4218-4231 (2007).
98. Sparvero, L., Asafu-Adjei, D., Kang, R., Tang, D., Amin, M., Im, J., Rutledge, R. Lin, B., Amoscato, A.A., Jeh, H.J. & Lotze, M.T. RAGE (receptor for advanced glycation end products), RAGE ligands, and their role in cancer and inflammation. *J. Transl. Med.* **7:17** (2009).
99. Fuentes, M.K., Nigavekar, S.S., Arumugam, T., Logsdon, C.D., Schmidt, A.M., Park, J.C. & Huang, E.H. RAGE activation by S100P in colon cancer stimulates growth, migration and cell signalling pathways. *Dis. Colon Rectum* **50,** 1230-1240 (2007).
100. Schmidt, A. M., Vianna, M., Gerlach, M., Brett, J., Ryan, K., Kao, C., Esposito, H., Hegarty, W., & Hurley, M. Isolation and characterisation of two binding proteins for advanced glycosylation end products from bovine lung which are present on the endothelial surface. *J. Biol. Chem.* **262,** 14987-14997 (1992).
101. Neeper, M., Schmidt, A. M., Brett, J., Yan, S. D., Wang, F., Pan, Y. C., Elliston, K., Stern, D. & Shaw, A. Cloning and expression of a cell surface receptor for advanced glycosylation end products of proteins. *J. Biol. Chem.* **267,** 14998-15004 (1992).

102. Kiselyov, V.V., Berezin, V., Maar, T.E., Soroka, V., Edvardsen, A., Schousboe, E. & Bock, E. The first immunoglobulin - like neural cell adhesion molecule (NCAM) domain is involved in double reciprocal interaction with the second immunoglobulin like NCAM domain and in heparin binding. *J. Biol. Chem.* **272,** 10125-10134 (1997).
103. Barclay, A. N. Membrane proteins with immunoglobulin - like domains - a master superfamily of interaction molecules, *Semin. Immunol.* **15,** 215-223 (2003).
104. Liliensiek, B., Weigand, M .A., Bierhaus, A., Nicklas, W., Kasper, M., Hofer, S., Plachky, J., Grone, H. J., Kurschus, F. C., Schmidt, A. M., Yan, S. D., Martin, E., Schleicher, E., Stern, D. M., Hammerling, G. G., Nawroth, P. P. & Arnold, B. Receptor for advanced glycation end products (RAGE) regulates sepsis but not the adaptive immune response. *J. Clin. Invest.* **113,** 1641-1650 (2004).
105. Schmidt, A. M., Yan, S. D., Yan, S. F. & Stern, D.M. The biology of the receptor for advanced glycation end products and its ligands. *Biochim. Biophys. Acta* **1498,** 99-111 (2000).
106. Xie, J., Reverdatto, S., Frolov, A., Hoffmann, R., Burz, D. S., Shektman, A. Structural basis for pattern recognition by the receptor for advanced glycation end products (RAGE). *J. Biol. Chem.* **283,** 27255-27269 (2008).
107. Srikrishna, G., Huttunen, H. J., Johansson, L., Weigle, B., Yamaguchi, Y., Rauyala, H. & Freeze, H. H. N-Glycans on the receptor for advanced glycation end products influence amphoterin binding and neurite outgrowth. *J. Neurochem.* **80,** 998-1008 (2002).
108. Leclerc, E., Fritz, G., Vetter, S.W. & Heizmann, C.W. Binding of S100 proteins to RAGE: an update. *Biochim. Biophys. Acta* **1793,** 993-1007 (2009).
109. Hori, O., Brett, J., Slattery, T., Cao, R., Zhang, J., Chen, J. X., Nagashima, M., Lundh, E. R., Vijay, S. & Nitecki, D. The receptor for advanced glycation end products (RAGE) is a cellular binding site for amphoterin. Mediation of neurite outgrowth and co-expression of RAGE and amphoterin in the developing nervous system. *J. Biol. Chem.* **270,** 25752-25761 (1995).
110. Deane, R., Yan, S. D., Submamaryan, R. K., LaRue, B., Jovanovic, S., Hogg, E., Welch, D., Manness, L., Lin, C., Yu, J., Zhu, H., Ghiso, J., Frangione, B., Stern, A., Schmidt, A.M., Armstong, D.L., Arnold, B.,,Liliensiek, B., Nawroth, P., Hofman, F., Kindy, M., Stern, D. & Zlokovic, B. RAGE mediates amyloid-beta peptide transport across the blood-brain barrier and accumulation in brain. *Nat. Med.* **9,** 907-913 (2003).

111. Wilton, R., Yousef, M. A., Saxena, P., Szpunar, M. & Stevens, F.J. Expression and purification of recombinant human receptor for advanced glycation end products in Escherichia coli. *Protein Expression purify.* **47,** 25-35 (2006).

112. Sousa, M. M., Yan, S. D., Stern, D. & Saraiva, M. J. Interaction of the receptor for advanced glycation end products (RAGE) with transthyretin triggers nuclear transcription factor kB (NF-kB) activation. *Lab. Invest.* **80,** 1101-1110 (2000).

113. Hofmann, M. A., Drury, S., Fu, C., Qu, W., Taguchi, A., Lu, Y., Avila, C., Kambham, N., Bierhaus, A.M., Nawroth, P., Neurath, M.F., Slattery, T., Beach, D., McClary, J., Nagashima, M., Morser, J., Stern, D. & Schmidt, A.M. RAGE mediated novel proinflammatory axis: a central cell surface receptor for S100/calgranulin polypeptides. *Cell* **97,** 889-901 (1997).

114. Chavakis, T., Bierhaus, A., Al-Fakhri, N., Schneider, S., Witte, T., Linn, M., Nagashima, J., Morser, B., Arnold, K.T., Preissner, P.P. & Nawroth, P.P. The pattern recognition receptor (RAGE) is a counter-receptor for leukocyte integrins: a novel pathway for inflammatory cell recruitment. *J. Exp. Med.* **198,** 1507-1515 (2003).

115. Schmidt, A. M., Hofmann, M., Taguchi, A., Yan, S.D. & Stern, D. M. RAGE : a multiligand receptor contributing to the cellular response in diabetic vasculopathy and inflammation. *Semin. Thromb. Hemost.* **26,** 485-493 (2000).

116. Brett, J., Schmidt, A. M., Yan, S. D., Zou, Y. S., Weidmann, E., Pinsky, D., Nowygrod, R., Neeper, M., Przysiecki, C & Shaw, A. Survey of the distribution of a newly characterised receptor for advanced glycation end products in tissues. *Am. J. Pathol.* **143,** 1699-1712 (1993).

117. Ostendorp, T., Leclerc, E., Galichet, A., Koch, M., Demling, N., Weigle, B., Heizmann, C.W., Kroneck, P.M.H. & Fritz, G. Structural and functional insights into RAGE activation by multimeric S100B. *EMBO J.* **26,** 3868-3878 (2007).

118. Xie, J., Burz, D.S., He, W., Bronstein, I.B., Lednev, I. & Shektman, A. Hexameric calgranulin C (S100A12) binds to the receptor for advanced glycation end products (RAGE) using symmetric hydrophobic target binding patches. *J. Biol. Chem.* **282(6),** 4218-4231 (2007).

119. Chin, D.H. & Means, A.R. Calmodulin: a prototypical calcium sensor. *Trends in Cell Biol.* **10(8),** 322-328 (2000).

120. Makalowski, W., Zhang, J. H. & Boguski, M. S. Comparative analysis of 1196 orthologous mouse and human full-length mRNA sequences. *Genome Res.* **6,** 846-857 (1996).

121. Freidberg, F. & Taliaferro, L. Calmodulin genes in zebrafish (revisited). *Mol. Biol. Rep.* **32,** 55-60 (2005).

122. Freidberg, F. & Rhoads, A. R. Sequence homology of the 3' untranslated region of calmodulin III in mammals. *Mol. Biol.Rep.* **28,** 27-30 (2001).

123. Copley, R. R., Schultz, J., Ponting, C. P. & Bork, P. Protein families in multicellular organisms. *Curr. Opin. Struct. Biol.* **9,** 408-415 (1999).

124. Bhattacharya, S., Bunick, C. G., Chazin, W. J. Target selectivity in calcium-binding proteins. *Biochim. Biophys. Acta* **1742,** 69-79 (2004).

125. Grabarek, Z. Structural basis for the diversity in EF-hand calcium binding proteins. *J. Mol. Biol.* **359,** 509-525.

126. Bahler, M., Rhoads, A. Calmodulin signalling via the IQ motif. *FEBS Letters* **513,** 107-113 (2002).

127. Jurado, L.A., Chockalingam, P.S. & Jarett, H. W. Apocalmodulin. *Physiol. Rev* . **79,** 661-682 (2009).

128. Valeyev, N. V., Heslon-Harrison, P., Postlethwaite, I., Kotov, N.V. & Bates, D.G. Multiple calcium binding sites make calmodulin functional. *Molecular Biosystems* **4,** 66-73 (2008).

129. Schaub, M. C., Heizmann, C. W. Calcium, calmodulin, troponin, S100 proteins: from myocardial basics to therapeutic strategies. *Biochem. Biophys. Res. Commun.* **369,** 247-264 (2008).

2. Materials and Methods

2.1 Cloning of S100 proteins

During the period of my PhD, I have cloned the following S100 proteins: S100A1, S100A2, S100A4, S100A5, S100A12, S100A14, S100P, S100Z. The cDNAs of all the S100 proteins were purchased from Imagenes. S100A1, A2, A5, A12, A14 and Z were cloned in the vector pET21a using the classical cloning method (using restriction enzymes). S00A4 was cloned in pET DEST 42 and S100P in pETG-30A using Gateway technology.

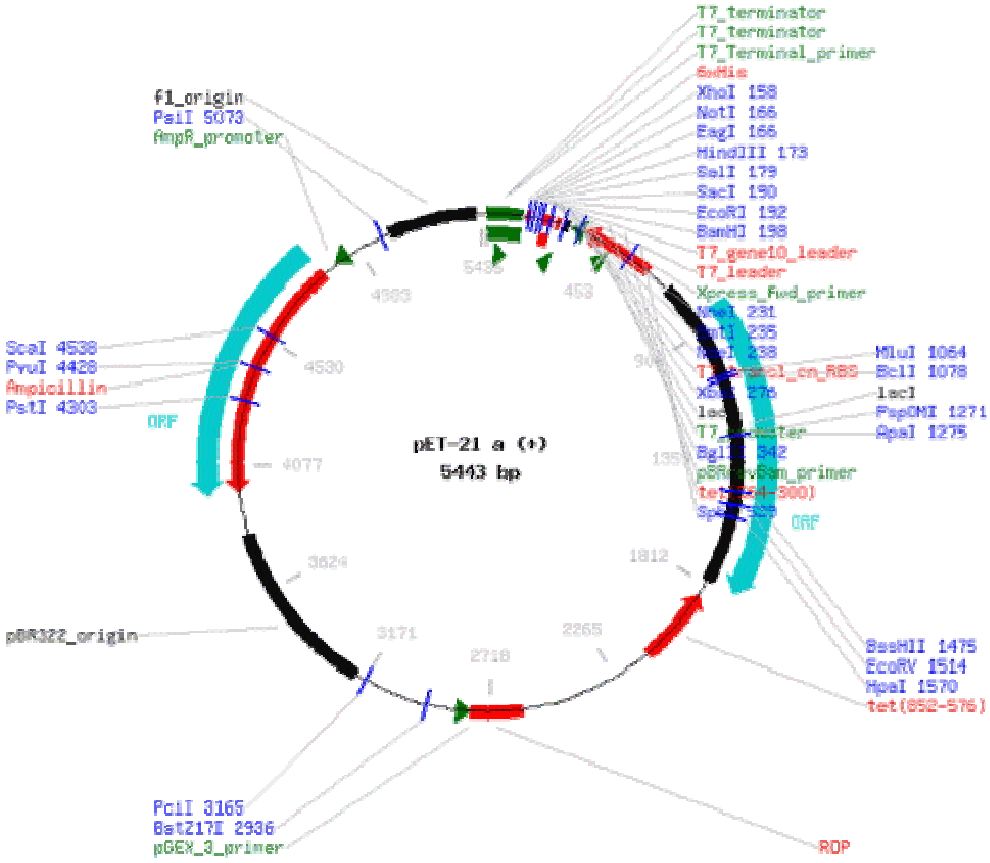


Figure 1: Map of pET21a vector

Primer designing

The NEB cutter tool was used to find out which restriction enzymes did not cleave the gene sequences of S100 proteins. None of the S100 genes contain the recognition site of *Nde*1. With the exception of S100Z, the other S100 genes also lack the *Xho*1 recognition sequence. As such, primers were designed for all S100 proteins (excepting S100Z) with *Nde*1 (CATATG) and *Xho*1 (CTCGAG) cleavage sites at the 5' ends of the forward and reverse complementary sequences respectively. For S100Z, *BamH*1 recognition sequence was introduced at the 5' end of the reverse primer. The recognition sequences were preceded with a few (4-5) bases for efficient cleavage by the restriction enzymes. A stop codon was introduced in the reverse primer after the restriction site to eliminate the C-terminal histidine tag in the vector.

For S100A4 and S100P which were cloned by gateway technology, the sequence CACCATG was introduced at the 5'end of the complementary forward primers. For S100A4, two stop codons were introduced at the 5' end of the complementary reverse primer to eliminate the C-terminal histidine tag of the vector.

PCR amplification

The gene sequences were amplified from the cDNA clone using P*fx*50^{TM} DNA polymerase from Invitrogen. It is a fusion enzyme consisting of recombinant DNA polymerase from the archaean *Thermococcus zilligi* fused to an accessory protein. The highly thermostable polymerase possesses a proofreading 3'-5' exonuclease activity, while the accessory protein stabilises the primer-template complexes in PCR. P*fx*50^{TM} offers 50 times better fidelity than TAQ DNA polymerase coupled with high specificity and an extremely fast elongation rate (15 seconds per kb). In addition, the fusion enzyme has an intrinsic hot-start capability for room temperature reaction assembly. It produces blunt end PCR products. PCR amplification was performed in a reaction volume of 50µl using 50 ng of template DNA. The cycling conditions used are as follows:

Initial denaturation at 94°C for 2 mins.

35 cycles of:

Denaturation: 94°C for 15 seconds

Annealing: 60-68°C (Tm of primers minus 2°) for 30 seconds

Extension: 68°C for 60 seconds

Final extension: 68°C for 5 minutes

The PCR amplified genes (270-320) base pairs were subsequently purified from the template, unreacted primers and enzyme with the PCR purification kit from Invitrogen. In other cases, the PCR product was resolved on a 1.6% low melting point (LMP) agarose gel using 1X TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA). The band was excised from the gel and purified by using the commercial gel extraction kit of Invitrogen.

Generation of Entry Clone

For S100A1, A2, A5, A12, A14 and Z the entry clone was generated using the CloneJet^{TM} PCR Cloning kit from Fermentas. It is an advanced positive selection system for high efficiency cloning of PCR products generated with any thermostable DNA polymerase. Any blunt or sticky end DNA fragment can be cloned and the kit is ideal for cloning phosphorylated or non-phosphorylated DNA fragments. The kit contains a linearised cloning vector pJET1.2/blunt which accepts inserts from 6 bp to 10 kb. The 5' ends of the vector cloning site contain phosphoryl groups. Therefore, phosphorylation of PCR primers is not required.

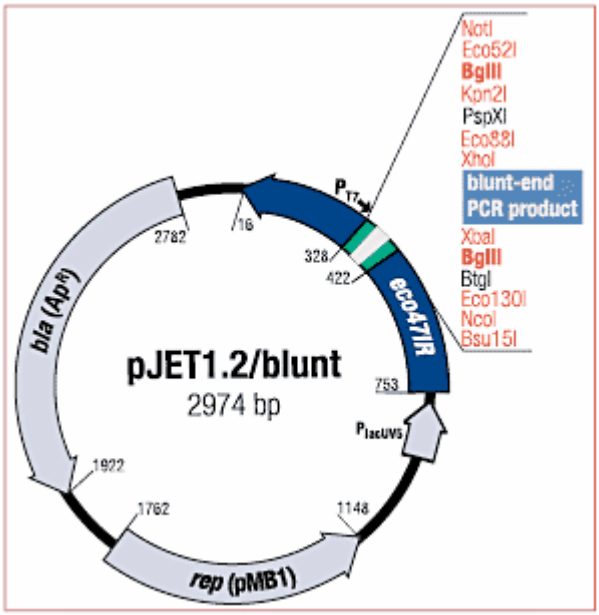
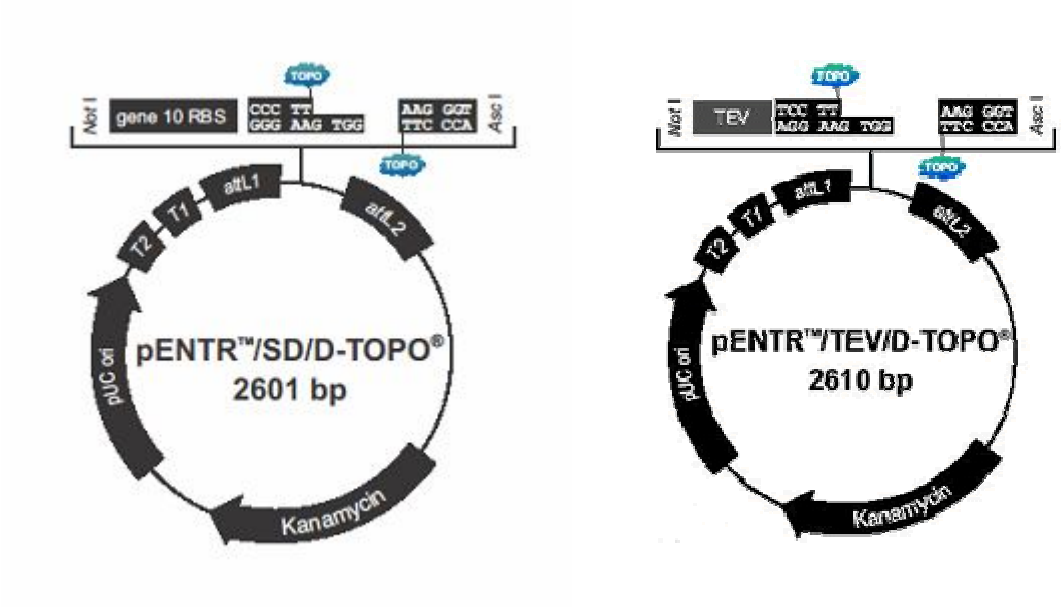


Figure 2: Map of pJET 1.2 vector

The pJET1.2 vector contains a lethal gene which is disrupted by insertion of a DNA insert into the cloning site. The recircularised vector expresses a lethal restriction enzyme after transformation and is not propagated. As a result, only recombinant clones containing the insert appear on culture plates. The vector is ampicillin resistant.

The blunt end PCR products generated by P*fx*50 were added to pJET1.2 in a molar ratio of 3:1 in presence of the reaction buffer and T4 DNA ligase and incubated at 22°C for 30 minutes. 5µl of the reaction mixture was used to transform 50µl of competent DH5α cells and plated on agar ampicillin plates and incubated overnight at 37°C.

For S100A4 and S100P the entry clones were generated by Topo cloning reaction in pENTR/SD/D-TOPO (for native protein) and pENTR/TEV/D-TOPO (for N-terminal fusion tags with TEV cleavage sites) respectively, using the directional Topo Cloning kit from Invitrogen.



S100A4 **S100P**

Figure 3: Maps of Gateway entry vectors pENTR/SD/D-TOPO (for native protein) and pENTR/TEV/D-TOPO (for N-terminal fusion proteins with TEV cleavage site

The blunt-ended PCR products were incubated in the topo cloning reaction mixture with the linearised vectors in 2:1 molar ratio for 30 minutes at room temperature for ligation. The vectors contain a single strand GTGG overhang on the 5' end and a blunt end on the 3' end. The four nucleotide overhang invades the double strand DNA of the PCR product and anneals to the CACC sequence in the 5' primer. Topoisomerase then ligates the PCR product in the correct orientation.

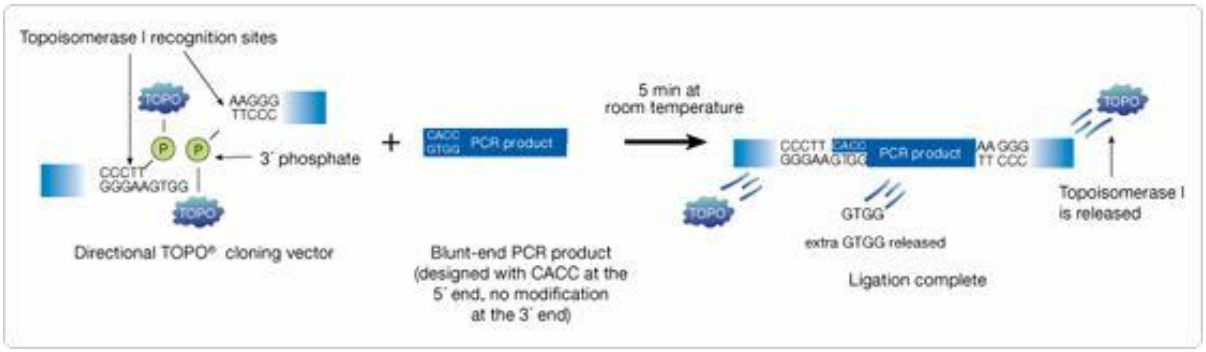


Figure 4: Schematic representation of Directional TOPO Cloning Reaction

The pENTR vectors contain kanamycin resistance gene for selection in E.coli. Therefore, the ligation mix was transformed into E. coli and spread on kanamycin plates.

Screening for Positive Constructs

For each S100 construct, a total of 10 colonies were screened by PCR amplification. To verify the correct orientation of the inserts, a mixture of vector specific forward primer and insert specific reverse primers were used. The PCR amplificates were checked on 1.6% agarose gel. The positive clones were purified and further verified by sequencing reactions.

Generation of Expression Clone

After sequencing, one positive construct was chosen for each S100 protein and proceeded with to create the expression clone.

Preparation of Insert: For all the proteins except S100A4 and S100P (for which gateway cloning technology was used), 1µg of the recombinant pJET1.2 plasmids with the different S100 genes were digested with the corresponding restriction enzymes whose recognition sites were introduced at the ends of the genes during PCR amplification. S100A1, A2, A5, A12, A14 were digested with *Nde*1 and *Xho*1 while S100Z was digested with *Nde*1 and *BamH*1. The recombinant plasmids were incubated with Fast Digest restriction enzymes from Fermentas for 1 hour at 37°C to excise the digested S100 genes from the plasmids. Restriction enzymes were inactivated by treating the reaction mixture at 80°C for 5 minutes. The entire reaction volume was loaded on a 1.6% agarose gel. The digested bands of S100 genes were excised and purified by the gel extraction kit from Invitrogen.

Preparation of Vector: 1µg of pET 21a was digested with *Nde*1 and *Xho*1 (for S100A1, A2, A5, A12, A14) or *Nde*1 and *BamH*1 (S100Z) by incubating the plasmid with Fast Digest restriction enzymes for 1hour at 37°C. The restriction enzyme was heat

inactivated at 80°C for 5 minutes. The linearised vector was treated with calf intestinal alkaline phosphatase (CIP) for 1 hour at 37°C to dephosphorylate its 5'ends and thereby minimise the chances of recircularisation. CIP was heat inactivated at 65°C for 20 minutes. The reaction mixture was loaded on a 0.8% gel and the linearised vector was excised and purified by gel extraction kit (Invitrogen).

Ligation: Ligation was performed using either the Rapid Ligation kit or the T4 DNA ligase from Fermentas. The digested insert and vector were mixed in the ratio 0f 3:1. For quick ligation, the reaction was incubated at 22°C for 1 hour while for T4 DNA ligase incubation was done at 22°C for 2 hours or 4°C overnight. 5µl of the ligation reactions were transformed into E. coli DH5α competent cells and spread on ampicillin plates for overnight incubation.

LR reaction: For S100A4 and S100P, which were cloned in gateway entry vectors, the expression clone was generated by transferring the S100 genes in destination vectors by the LR reaction. The LR reaction mix containing entry clone and destination vector was incubated at 25°C for 60 minutes. To terminate the reaction, proteinase K solution was added and the sample was incubated at 37°C for 10 minutes.

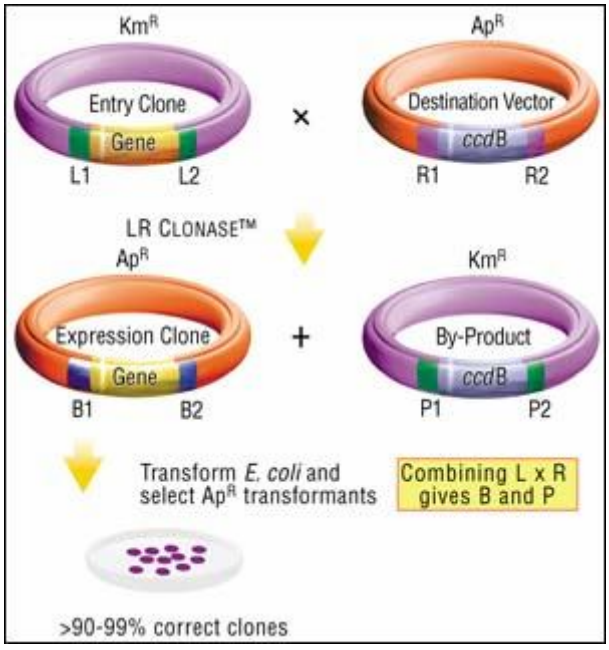


Figure 5: Schematic Representation of LR Reaction

The S100A4 gene was transferred to the expression vector pET DEST 42 (without fusion tags) while S100P was transferred to the vector pETG-30A with a N-terminal GST and

hexahistidine tag. 1µl of the reaction was transformed into chemically competent DH5α cells, spread on LB ampicillin plates and incubated overnight at 37°C.

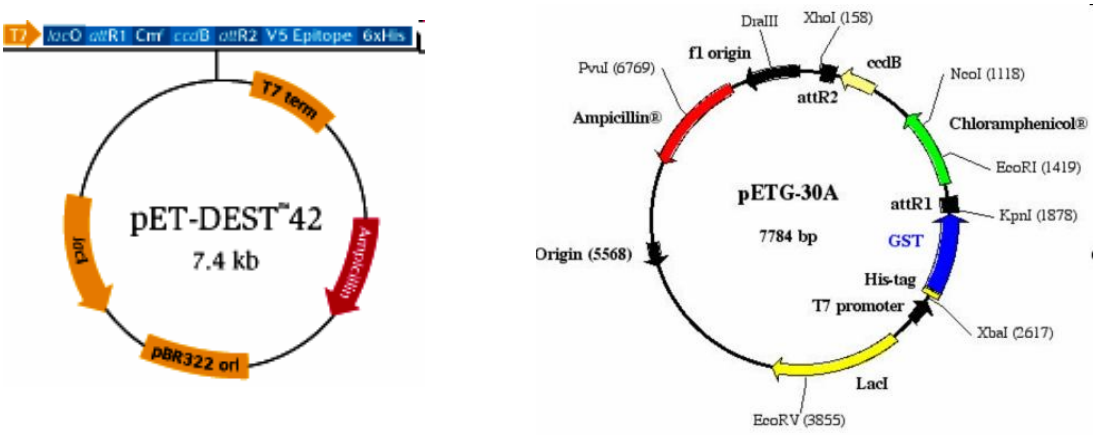


Figure 6: pET-DEST 42-destination vector for S100A4 without N-terminal fusion tag; pETG-30A-destination vector for S100P with N-terminal GST and histidine tag.

Screening for Positive Constructs

To screen for the positive colonies, 10 colonies of each S100 protein were analysed by colony PCR. To verify the orientation of the insert in the vector, a gene specific forward primer and vector specific reverse primer (T7 terminator) was used in the PCR reaction. The amplificates were resolved in a 1.6% agarose gel. One of the the positive colonies of each protein were chosen for expression tests.

2.2 Expression of S100 proteins

The recombinant plasmids carrying different S100 genes were transformed into different E. coli expression strains: BL21(DE3), Gold, pLys S, Codon Plus DE3) RIL. Protein expression was checked in each of them with 5 ml cultures. The strain with optimal expression was chosen for further tests including varying conditions of inducer concentration, induction temperature and length of induction. The conditions checked for expression are as follows:

IPTG concentration: 0.4 mM, 0.5 mM, 0.7 mM, 1.0 mM.

Induction temperature: 37°, 30°, 25°, 20°, 17°C.

Induction period: Every 1 hour after induction upto 6 hours and overnight (12-16 hours).

For each S100 protein, expression was checked under the above conditions. Samples of soluble and insoluble cellular fractions were analysed on 15% SDS-PAGE. The cellular fractions which contain the protein and optimal expression conditions were determined for large scale protein expression. The optimised expression conditions for all the S100 proteins have been summarised in the table below.

S100 Protein	Plasmid	E. coli strain	IPTG Concentration	Induction Temperature	Induction Period
S100A1	pET21a	Gold	1.0 mM	37°C	ON
S100A2	pET21a	Gold	0.5 mM	37°C	ON
S100A4	pET DEST 42	pLys S	0.5 mM	25°C	ON
S100A5	pET21a	Gold	0.5 mM	25°C	ON
S100A12	pET21a	Gold	1.0 mM	37°C	ON
S100P	pETG-30A	Gold	1.0 mM	25°C	ON
S100A14	pET21a	Codon Plus	0.5 mM	37°C	ON
S100Z	pET21a	Gold	0.5 mM	37°C	ON

2.3 Purification of S100 proteins

Protein purification was based on the intrinsic properties of S100 proteins:

a) Solubility of S100 proteins in 100% saturated ammonium sulphate solution at neutral pH.

b) Calcium binding to most S100 proteins exposes a patch of hydrophobic residues on the protein surface which otherwise remain buried in the apo form.

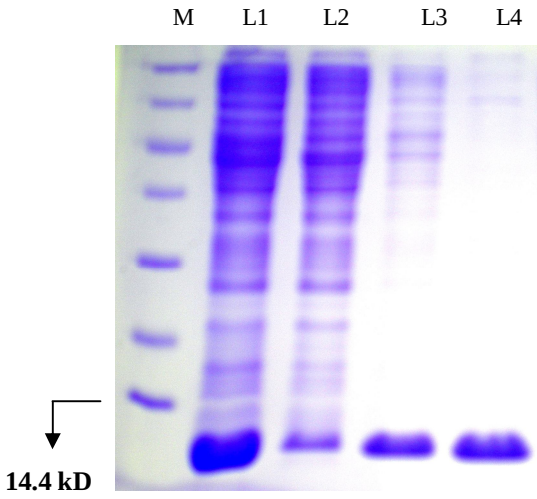
S100 A1, A2, A5 and A12 was purified as follows:

1) Bacterial cells pelleted from the culture were resuspended in lysis buffer 20 mM Tris, 1mM PMSF, 2 mM DTT, 2 mM DTT, pH 7.5. In case of S100A12, 100 mM NaCl was included in the buffer.
2) Cells were lysed by sonication and the suspension was ultracentrifuged at 40,000 rpm for 30 minutes.
3) S100 proteins were present in the soluble cellular fraction in all cases. So the clarified lysate was subjected to purification.

   For S100A5, 30% ammonium sulphate was added to the lysate at 4°C with constant stirring to precipitate out impurities. The suspension was kept in stirring condition for at least 2 hours to allow it to become homogeneous. After the stipulated period, it was centrifuged at 9000 rpm for 20 mins. The pellet of precipitated proteins was discarded and the supernatant containing S100A5 was collected for further purification.
4) To the lysate of A1, A2, A12 and the supernatant from step 3 of A5, 5 mM $CaCl_2$ was added to convert the S100 proteins to their holo form and thereby expose the hydrophobic amino acids at the surface. S100A12 precipitates out of the lysate when $CaCl_{2\ is}$ added to it directly. So A12 lysate was diluted with buffer containing $CaCl_2$ in the ratio of 1:5. After addition of $CaCl_2$ the suspension was clarified by centrifugation in all cases.

Chromatographic steps:

a) Hydrophobic Interaction Chromatography: A 5 ml Phenyl Sepharose high sub column (from Amersham Biosciences) was equilibrated with lysis buffer containing $CaCl_2$. The supernatant from step 4 was loaded on the column at a rate of 0.4-0.5 ml/min for optimal binding. Unbound proteins were washed out of the column with 12-15 column volumes of equilibration buffer. The bound S100 proteins were then eluted with 20 mM Tris, 10 mM EDTA, pH 7.5.



M: Marker, L1: Lysate, L2: Column flowthrough, L3: Wash, L4: Elution

Figure 7: Polyacrilamide gel showing fractions of S100 proteins from hydrophobic interaction chromatography.

b) Gel Filtration: The elution from the HIC column was concentrated and loaded on a Superdex 75 16/60 column. Proteins were eluted in the final buffer 30 mM MES, 100 mM NacCl, 5 mM DTT, pH 6.5.

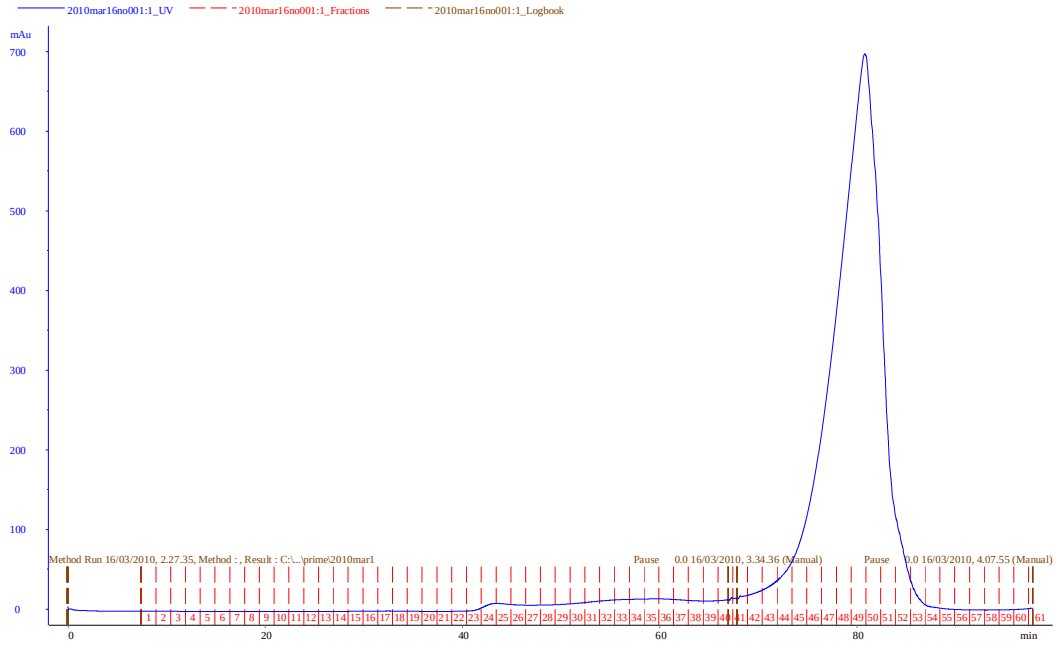


Figure 8: Chromatogram of gel filtration of S100 protein

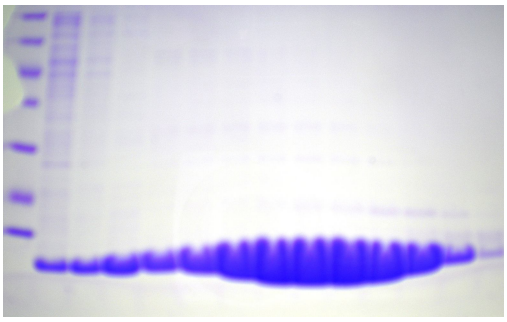


Figure 9: SDS PAGE of gel filtration fractions of S100 protein.

Purification of S100A4

a) Anion exchange: A HiTrap Q (stong anion exchange) column was equilibrated with 20 mM Tris, 100 mM NaCl, pH 8.0. The lysate was loaded on the column. S100A4 was recovered in the unbound fraction. The bound impurities were eluted from the column with 20 mM Tris, 1M NaCl, pH 8.0.

b) Cation exchange: The column flowthrough of anion exchange was buffer exchanged into 20 mM MES, 50 mM NaCl, pH 5.6. It was loaded on a Hi Trap SP (strong cation exchange) column equilibrated with the same buffer. Bound S100A4 was eluted with a salt gradient from 50-500 mM. The protein eluted at 300 mM salt concentration.

c) Gel filtration: The purest fractions of cation exchange were concentrated and loaded on a Superdex 75 16/60 column. S100A4 was eluted with 30 mM MES, 100 mM NaCl, 5 mM DTT, pH 6.5.

Purification and Refolding of S100A14

S100A14 is expressed in the inclusion bodies of E. coli. The protein was therefore extracted from the inclusion body in 20 mM Tris, 1 mM EDTA, 5 mM DTT, 8M urea, pH 8.0. The unfolded protein was purified by anion exchange using a Q sepharose column. S100A14 does not bind to Q column at pH 8. The bound impurities were eluted with the same buffer containing 800 mM NaCl. The unbound fractions containing the protein were refolded by dialysing it against a deacreasing urea gradient till the urea concentration reached zero. The folded protein was further purified by size exclusion chromatography and eluted in 30 mM MES, 100 mM NaCl, 5 mM DTT, pH 6.5.

Purification of S100P

S100P was expressed as a fusion protein with N-terminal GST and Histidine tag. As such it was purified by immobilised metal ion chromatography (IMAC). The bacterial pellet was resuspended in 1X phosphate buffer, 10 mM imidazole, pH 7.4 (binding buffer). Cells were lysed by sonication and the suspension ultracentrifuged. The lysate was purified as follows:

a) Lysate was loaded on a nickel sepharose column equilibrated with binding buffer.
b) Unbound proteins were washed out with several column volumes of binding buffer.
c) Impurities bound to the column were washed out with 1X phosphate buffer, 100 mM imidazole, pH 7.4.
d) The bound his-tagged S100P was eluted with 1X phosohate buffer, 500 mM imidazole, pH 7.4.

Tag cleavage

In pETG-30A a TEV protease recognition site is incorporated between the N-terminal tag and the recombinant gene. TEV protease is a highly site-specific cysteine protease that is found in Tobacco Etch Virus (TEV). The optimum recognition site for this enzyme is Glu-Asn-Leu-Tyr-Phe-Gln-(Gly/Ser) and cleavage occurs between the Gln or Gly/Ser residues. For cleaving the tag from S100P the elution from nickel column was first changed into TEV buffer (50 mM Tris-HCl , pH 8; 0.5 mM EDTA, 1 mM DTT). The fusion protein was incubated overnight with recombinant (histidine–tagged) TEV protease at room temperature.

Separation of cleaved S100P

The cleaved S100P was purified from the tag and TEV protease by passing the mixture over a nickel column. S100P was collected as column flow through. The bound tag and TEV protease were eluted with imidazole.

S100Z has been found to be expressed in the inclusion bodies. Its purification and refolding protocols are being optimised.

2.4 Paramagnetic Probe attachment to S100 proteins

Paramagnetic probes are important tools for studying protein structure and dynamics. They can be particularly useful in investigating protein-protein interactions, providing long range distance restraints to characterise the complex. To attach a paramagnetic metal to S100 proteins, the EF2 of several S100 proteins were mutated to tune their affinity for lanthanides and replace the calcium at the binding site with a lanthanide. Mutations were based on sequence alignment of the EF2 of these S100 proteins with S100G/calcindin D9K, whose EF2 has a strong affinity for lanthanides, so that the calcium at the binding site can be preferentially replaced by a paramagnetic metal.

```
                X*-Y*ZG#Ix**z
CALB9      :    DK-NGDGEVSFEE
S100A5     :    DK-NSDQEIDFKE
S100A1     :    DE-NGDGEVDFQE
S100P      :    DA-NGDAQVDFSE
S100A2     :    DE-NSDQEIDFKE
```

* = any residue; X = first calcium ligand; Y = second calcium ligand; Z = third calcium ligand; G = glycine; # = fourth calcium ligand, provided by a backbone carbonyl; I = isoleucine (although other aliphatic residues are also found at this position); x = fifth calcium ligand; z = sixth and seventh calcium ligands, provided by a bidentate glutamate or aspartate residue

Figure 10: Sequence alignment of EF-2 of S100 proteins with that of calbindin (S100G).

The mutations designed were as follows:

S100A5: Q65G, K70E

S100P : A67G, Q68E, S72E

S100A2: Q69E

The EF-2 of **S100A1** is strikingly similar to that of calbindin (differing in only two residues). So this protein was also expressed for paramagnetic studies.

Mutagenesis Protocol

The mutations listed above were performed using the Quick Change Site-Directed Mutagenesis kit from Stratagene as follows:

Mutagenic Primer Design: Both sense and antisense primers were designed in a way that they contained the desired mutation and annealed to the same sequence on opposite strands of the template plasmid. Primer lengths were kept typically between 30-35 bases. The desired mutation was in the middle of the primer with approximately 10-15 bases of correct sequence on both sides. Each primer terminated in G or C bases, had a GC content of about 40% and Tm $\geq$ 78°C. Tm was calculated using the following formula:

$$Tm = 81.5 + 0.41(\%GC) – 675/N - \%\ mismatch$$

where N is the primer length in bases, values of %GC and % mismatch are whole numbers.

Mutant Strand Synthesis: Recombinant plasmid DNA (containing the wild type gene) extracted from DH5α (dam^+ strain) was amplified by *Pfu Turbo* DNA polymerase using 125 ng of each mutagenic primer. Cycling conditions were determined depending upon the type of mutation (point / single amino acid change).

***Dpn* I digestion of template:** *Dpn* 1 endonuclease is specific for methylated and hemimethylated DNA. Parental plasmid DNA was selectively digested leaving the synthesised mutated DNA.

Transformation: The amplified mutant DNA was transformed into XL1-Blue supercompetent cells and incubated at 37° C for 16-18 hours with suitable antibiotic selection.

The mutations were verified by sequencing reactions.

The mutants were expressed and purified in the same way as the wild type proteins.

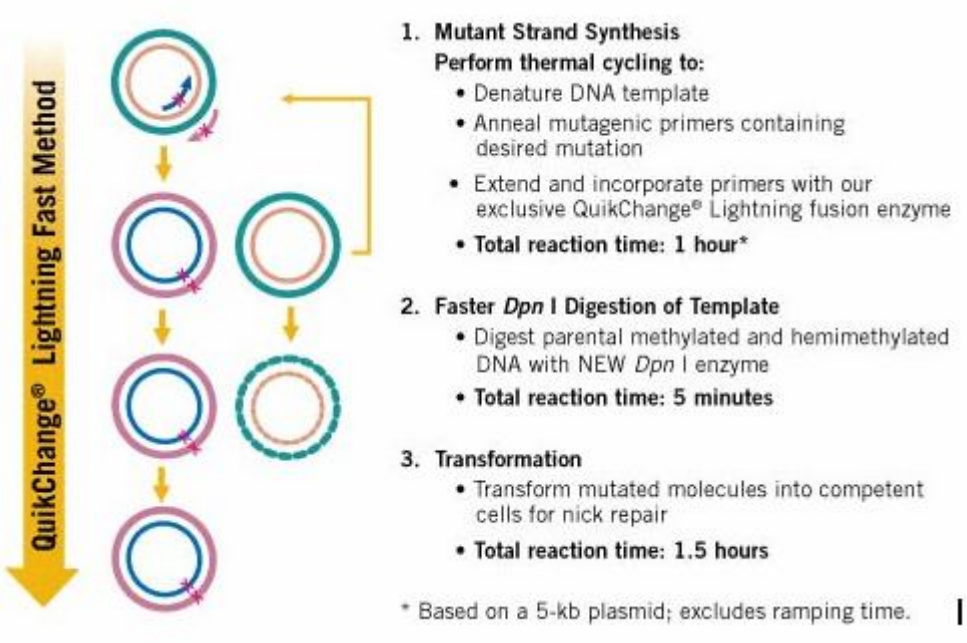


Figure 11: Schematic representation of site-directed mutagenesis procedure.

2.5 Mammalian Cell Culture and Experimental Design

Cell Culture: The colorectal adenocarcinoma cell line (SW480) and pancreatic carcinoma cell line (Panc-1) were obtained from the American Type Culture Collection. Cells were grown in high glucose Dublecco's modified Eagle's medium (DMEM) supplemented with 10% foetal bovine serum (FBS), 2 mM glutamine, penicillin (100U/ml), streptomycin (100U/ml) and maintained in a humified atmosphere with 5% CO_2 at 37°C.
Cells were cultured for at least two weeks before using them for experiments with regular splitting upon reaching confluency.
RAGE overexpression at the cell surface was verified for both the cell lines with western blotting.

Experimental Design

For all the experiments cells were seeded in 96-well plates ($1x10^3$ cells/well in a volume of 100 µl) and incubated for 24 hours to enable them to attach to the extracellular matrix at the bottom of the wells. After 24 hours, the medium and unattached cells were removed and serum free medium was added to each well (serum may interfere with protein-protein interactions and give misleading results). This was followed by recombinant protein treatment and incubation for specific time periods. The experiments performed are outlined below:

1. Cells were treated with different concentrations (0.1-1-10-100-1000 nM) of recombinant S100A1, A2, A4, A5, A12, S100P and incubated for 72 hours. The effect of each S100 treatment on cell proliferation was examined. The concentration of each S100 with maximal effect on cell proliferation was determined for both the cell lines.

2. Cells were treated with the optimal concentration of each S100 protein as determined from experiment 1. Proliferation was measured each 24 hour after treatment upto 96 hours. The growth curves of SW480 and Panc-1 in response to treatment with the optimal concentration of each S100 protein were plotted upto 96 hours after treatment.

3. The extracellular part of RAGE consists of 3 domains-V, C1 and C2. Different S100 proteins bind to separate RAGE domains to activate it. Competition experiments were performed where cells were treated with S100 proteins together the three domains of

RAGE separately. The effect on cell proliferation for each domain plus S100 was assayed. The effect on cell proliferation upon treatment with S100 proteins alone and S100 plus RAGE was compared.

Each experimental condition was set up in triplicates to minimise errors. Experiments were repeated at least thrice to verify the reliability of results.

Cell proliferation was measured with the CellTiter 96 AQueous One Solution cell proliferation assay (MTS; Promega) which is a colorimetric method for determining the number of viable cells. The CellTiter 96 AQueous One Solution Reagent contains a tetrazolium compound, MTS and an electron coupling regent, PES. PES has enhanced stability which allows it to be combined with MTS to form a stable solution. MTS is bioreduced by cells into a coloured formazan product that is soluble in tissue culture medium. This conversion is accomplished by NADPH or NADH produced by dehydrogenase enzymes in metabolically active or viable cells.
For measuring cell proliferation, 20 µl of the reagent was added to each well. The plates were incubated for 1.5 – 2 hours at 37°C in humified atmosphere with 5% CO_2. Absorbance was recorded at 490 nm with a 96-well plate reader.

2.6 Paramagnetic Tag attachment to Calmodulin

Two residues in the C-terminal domain of CaM, H107C and N111C were mutated to cysteines for attachment of the Caged Lanthanide NMR Probe-5 (CLaNP-5) which attaches to proteins by two disulphide bonds.

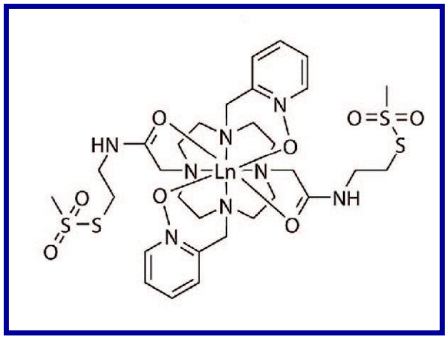


Figure 12: Caged Lanthanide NMR Probe-5

The advantages of this probe are:

A) Due to the presence of two pyridine-N-oxide ligating pendant arms, CLaNP-5 is symmetric.

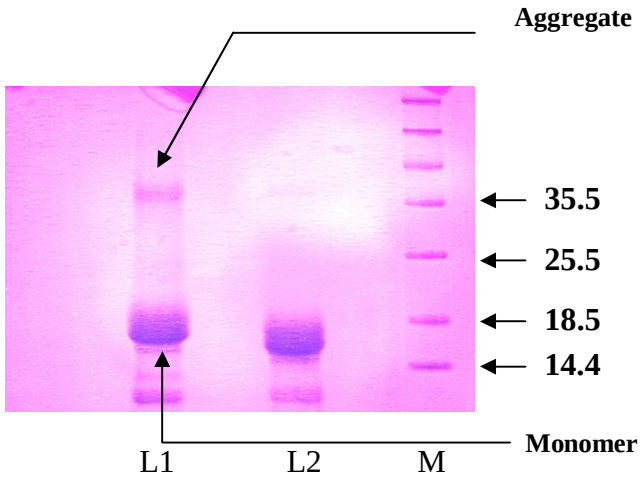
B) Two-point attachment of CLaNP-5 through disulphide bridges introduces large and well-defined paramagnetic effects by minimising the dynamics of the probe.

Protein Expression and Purification

The mutant CaM which is cloned in pET 16b has a His (10)-tag separated from the gene sequence by a factor Xa cleavage site. The protein was expressed in BL21 DE3 Gold cells and purified by a nickel sepharose column followed by tag cleavage. The native protein was then purified by a second metal affinity chromatography and gel filtration. For NMR experiments, the protein samples were prepared in 30 mM MES, 200 mM KCl, pH 6.8

Tag attachment Reaction

For attachment of Ln-CLaNP-5 to H107C/N111C CaM, all operations were performed in the glove box. The $4Ca^{2+}$- CaM mutant was reduced by incubating it with 5 mM DTT. Calcium was used in excess to avoid the possibility of exchange of calcium ions at the binding site with any unchelated lanthanide in the tag solution. DTT was removed under reducing conditions. The reduced protein was diluted to 20-30 µM and incubated with seven equivalents of Ln-CLaNP-5 in the glove box for 3-4 hrs at room temperature for the tag attachment to proceed to completion. . About 1/3 rd of the protein aggregated during the tag attachment reaction as observed in non-reducing SDS-PAGE.



M: marker (kD) L1: Tagged CaM (without DTT) L2: Untagged CaM (without DTT). **Figure 13:** Non-denaturing polyacrilamide gel showing the aggregation state of L2: calcium calmodulin and L1: calcium calmodulin reacted with the Ln-CLaNP-5.

The reacted sample was purified by a superdex 200 16/60 column to separate the monomer from aggregates. The protein was eluted in 30 mM MES, 200 mM KCl, 20 mM $CaCl_2$, pH 6.8.

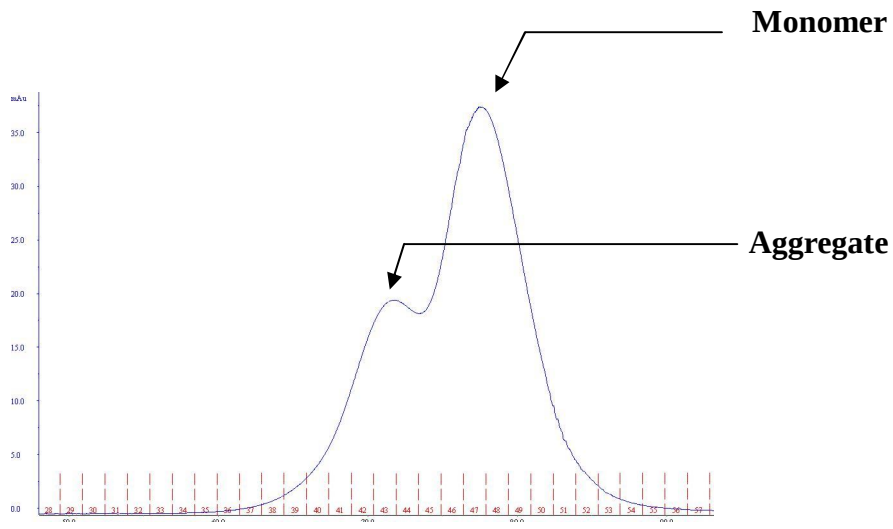


Figure 14: Chromatogram showing elution profile of CaM aggregate and monomer from superdex 200 column.

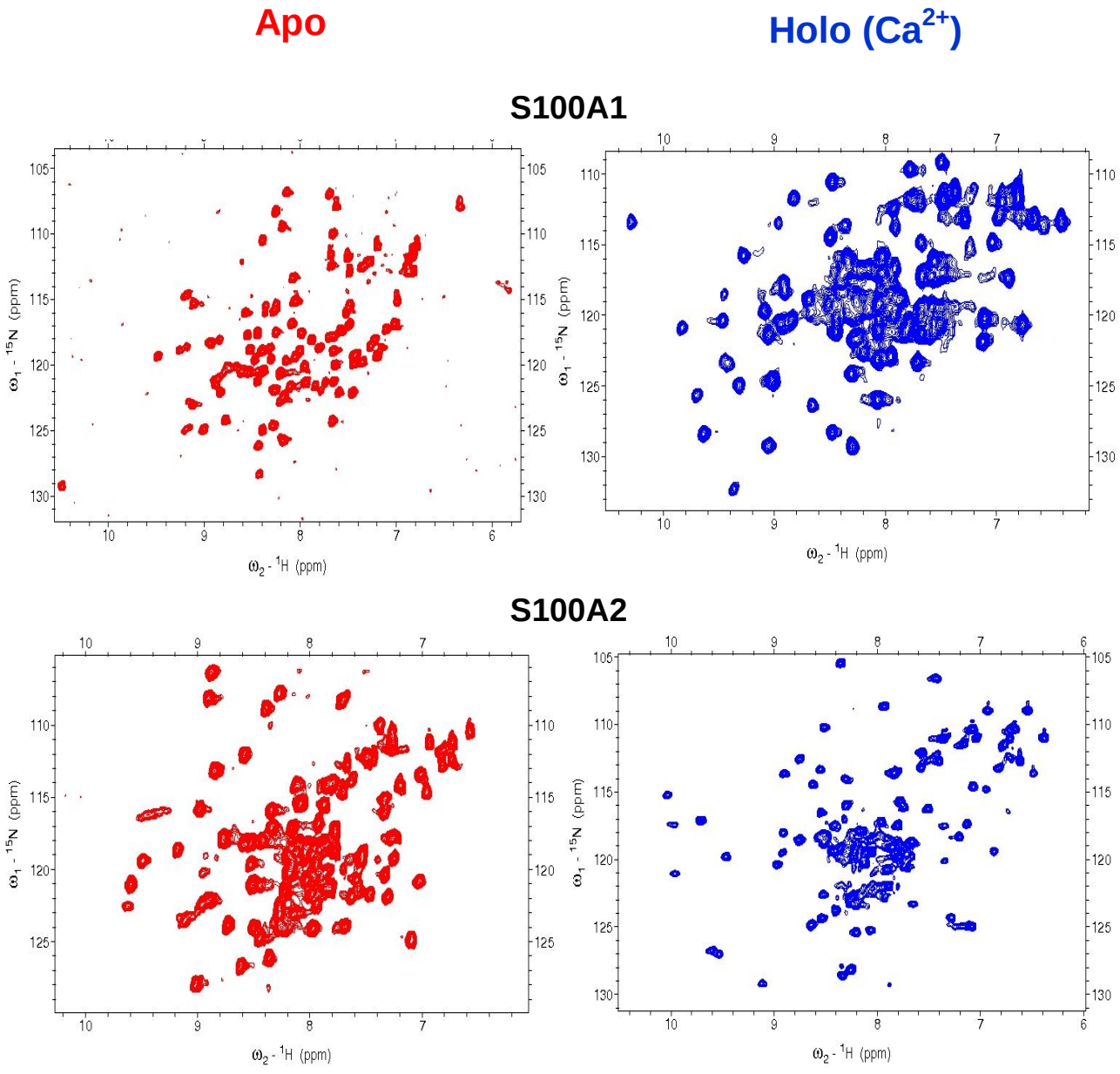
Tag attachment was verified by comparing the $^{1}H^{15}N$-HSQC of the unreacted CaM mutant with that of the reacted monomer.

3. Results

3.1 Structural characterisation of S100 proteins

Using the classical cloning method and pET expression system, a significant improvement in the expression and yield of S100 proteins were obtained in comparison to the Gateway cloning technology. S100 proteins have very low expression when cloned in gateway vectors. Highest expression of these proteins were found when they were cloned in the native state, i.e., without any fusion tags.

To check the folding and quality of the S100 proteins, 15N HSQC of the apo and Ca^{2+} titrated proteins were taken. Following are the HSQC spectra of apo- and holo- S100 proteins:



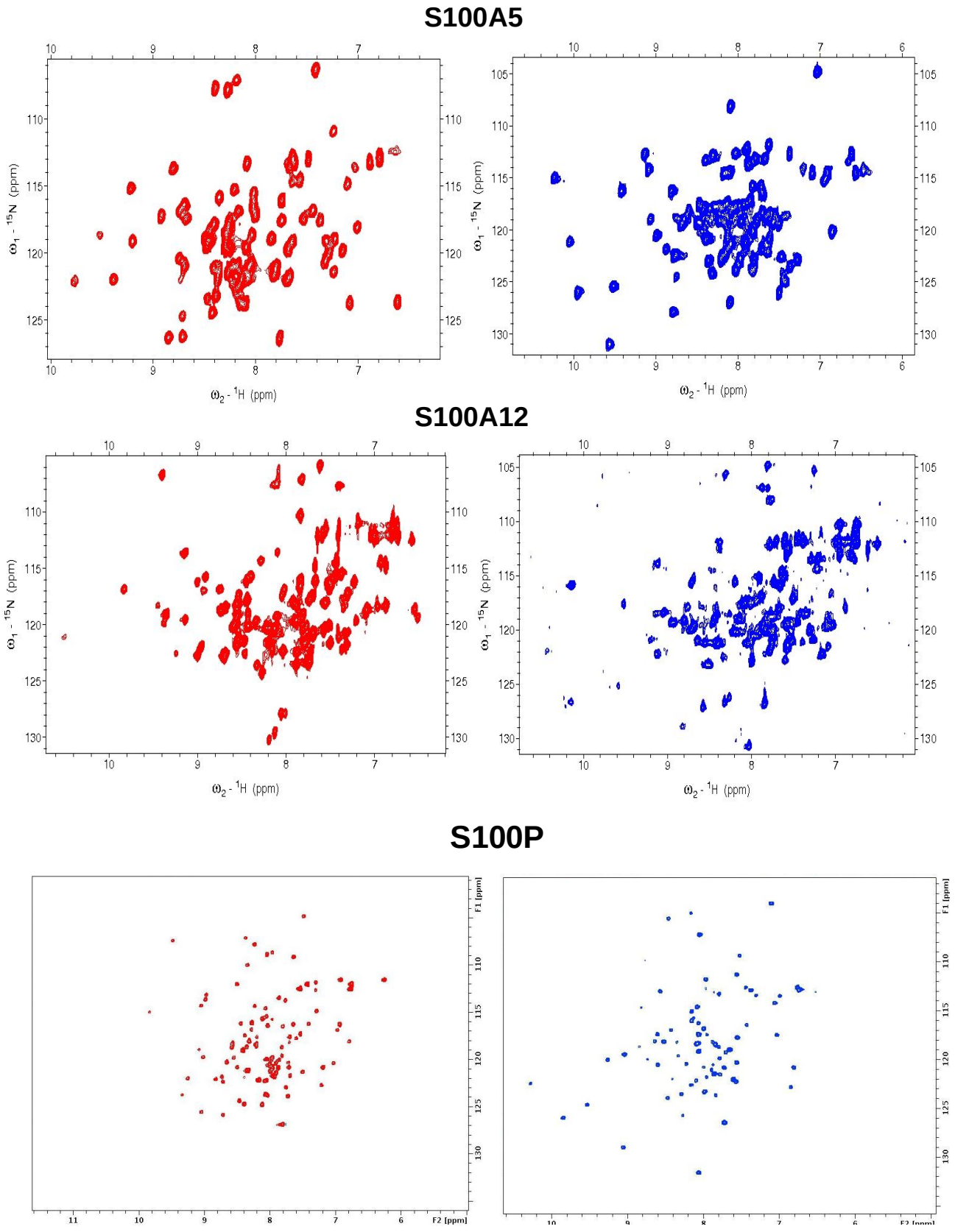


Figure 1: $^{1}H^{15}N$ HSQC spectra of apo and holo S100 proteins.

S100A14 which is expressed in the inclusion bodies was refolded and the apo spectrum of the protein shows that it is well folded.

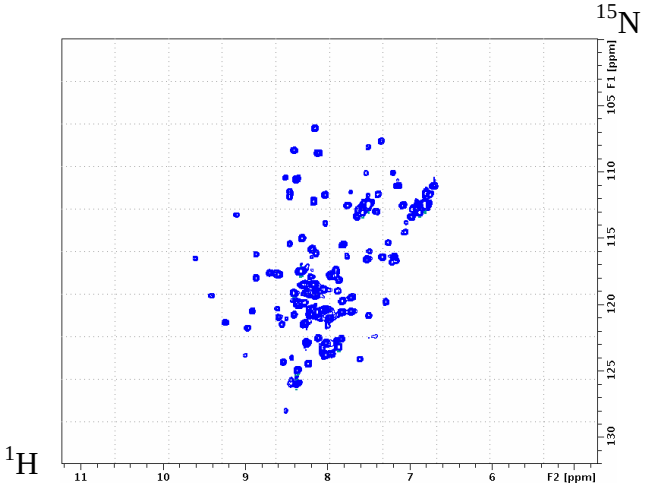


Figure 2: $^{1}H^{15}N$ HSQC spectra of apo S100A14

The above ^{15}N HSQC spectra were taken at 700 MHz at 298 K. Samples were prepared in 30 mM MES, 100 mM NaCl, 5 mM DTT, pH6.5. It is evident that all the above S100 proteins were folded and appropriate for use in NMR experiments.

S100A4 (apo)

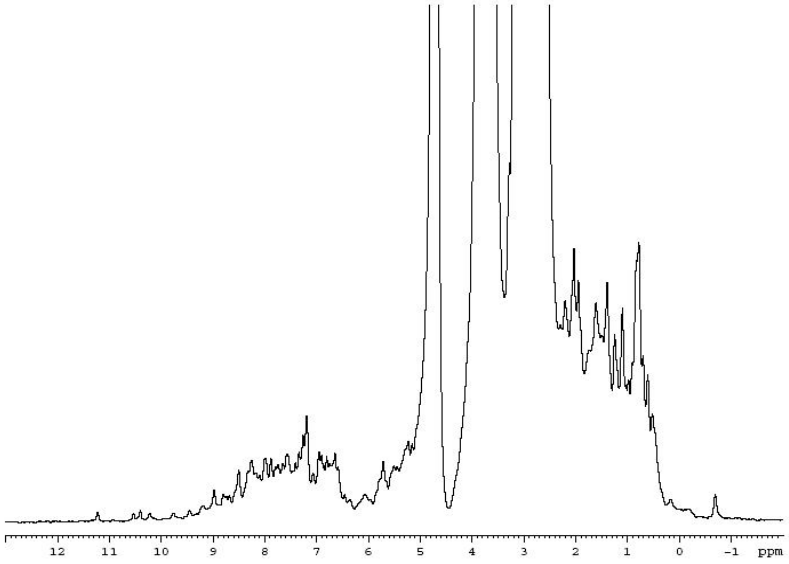


Figure 3: 1D spectrum of S100A4. Buffer: 30 mM MES, 100 mM NaCl, 5 mM DTT, pH 6.5. Temp: 298 K. Spectrum taken at 600 MHz.

S100A4 was expressed only for cell line studies. Hence only unlabelled sample of this protein was prepared.

One of the aims of this thesis was structural characterisation of S100 proteins. The structure of S100A5 was therefore solved by solution NMR.

3.2

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

*Bertini I, **Das Gupta S**, Hu X, Karavelas T, Luchinat C, Parigi G, Yuan J.* ***J Biol Inorg Chem*** *(2009) 14 : 1097-1107.*

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

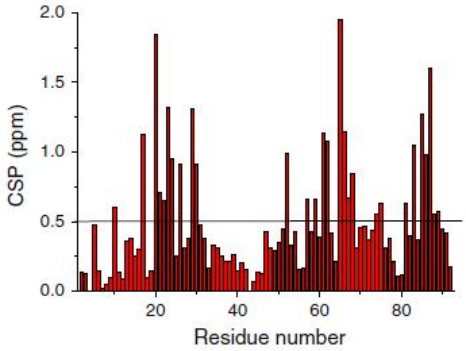


Fig. 1 Composite chemical shift perturbation (*CSP*), as defined in Eq. 1, of apo-S100A5 upon calcium(II) binding. The *horizontal line* indicates the average value

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).

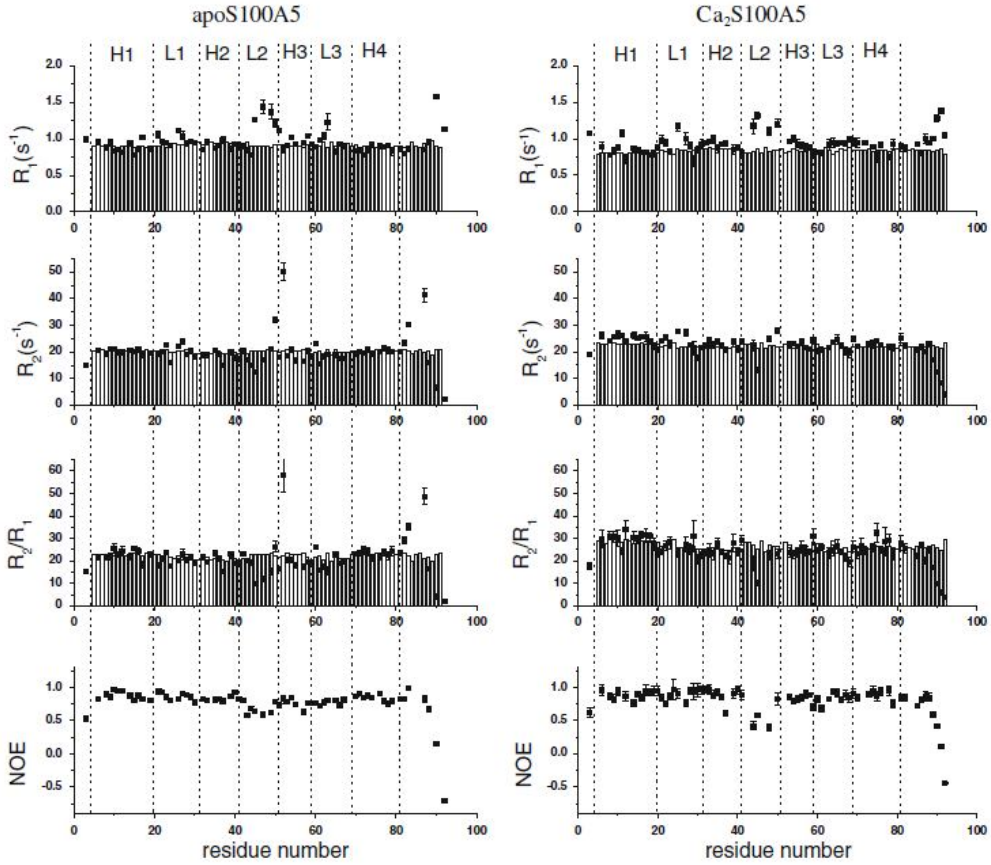


Fig. 2 Sequential plot of the experimental relaxation parameters of apo and Ca^{2+}-S100A5. The values calculated by HYDRONMR are shown as *bars*. *NOE* nuclear Overhauser effect

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 ^1H–^1H-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 ^1H–^{15}N-NOE being even smaller than in the apo form (the average ^1H–^{15}N-NOE in the hinge loop is 0.60 and 0.40 for the apo and calcium 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.

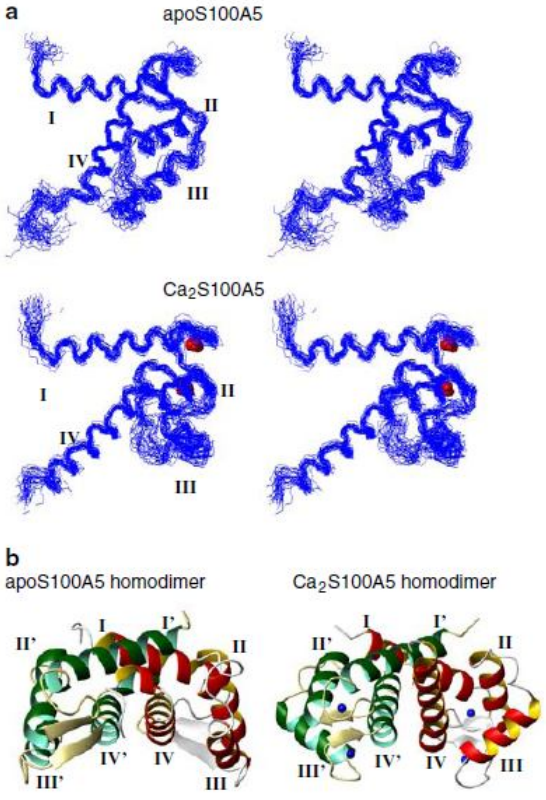


Fig. 3 Stereoview of the families of the solution structures of the S100A5 subunit in the apo and calcium(II)-loaded forms (**a**) and ribbon representation of the homodimer mean structures (**b**) obtained after restrained energy minimization

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

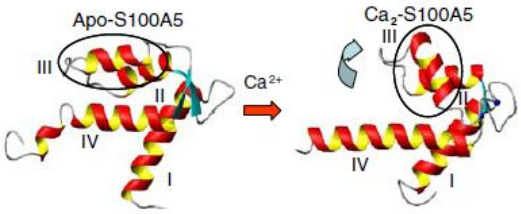


Fig. 4 Major structural differences upon calcium binding: the different angle between helices III (enclosed in *circles*) and IV, and the longer α-helical structure of helix IV

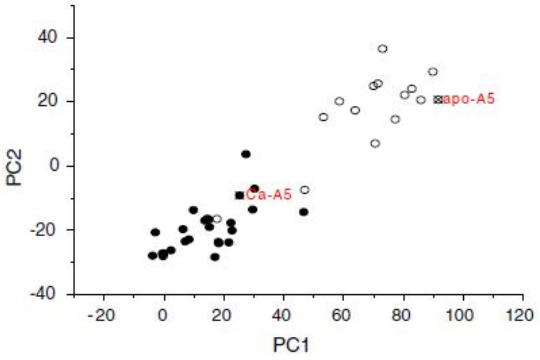


Fig. 5 Principal component plot for the S100 proteins derived from principal components analysis of the six interhelix angles [21]. Apoproteins (S100A1, S100A2, S100A3, S100A4, S100A5, S100A6, S100A10, S100A11, S100A13, S100B, calbindin D_{9k}) are indicated with *open circles* and calcium-loaded proteins (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

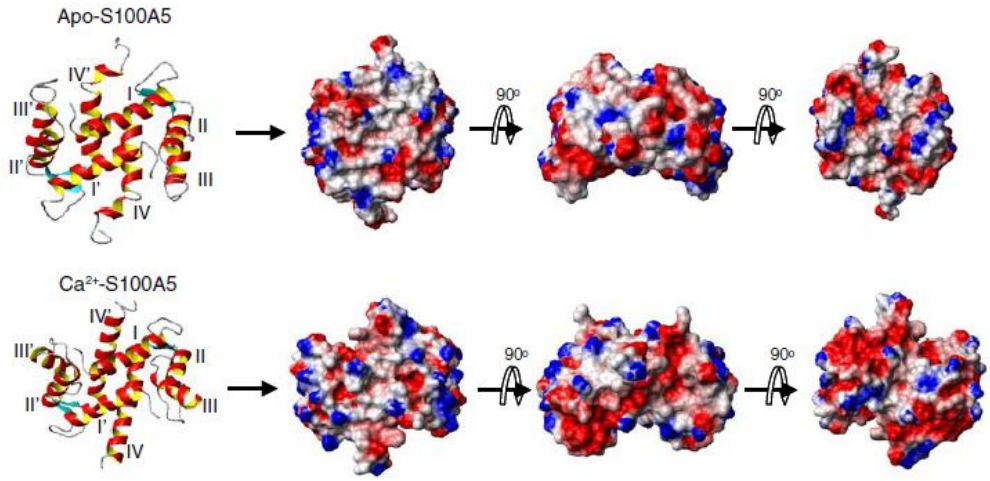


Fig. 6 Electrostatic surface representation of the S100A5 dimer

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

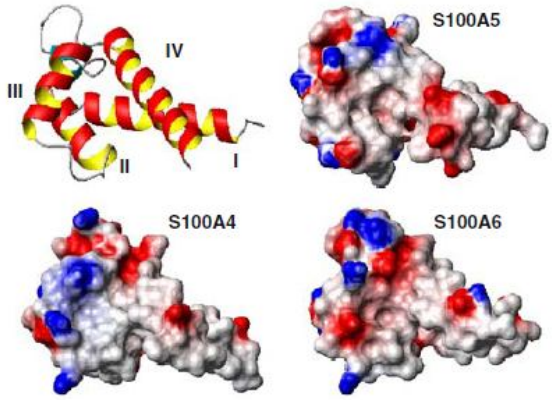


Fig. 7 Electrostatic surface representation of the reported ligand binding area typical of S100 proteins, for S100A4 (2Q91), S100A5 (present work), and S100A6 (1K9K)

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

Supplementary Material. Table S1. Interhelical angles and main principal components for S100 proteins, whose structures were not available at the time of the compilation of the list of S100 proteins published in the Supporting Information of Babini E, Bertini I, Capozzi F, Luchinat C, Quattrone A, Turano M (2005) J Proteome Res 4:1961-1971

PDB	Calcium state – peptide binding	Protein name	Selected residues for Helix I Helix II Helix III Helix IV	Inter-helix angle I / II I / III I / IV II / III II / IV III / IV	Principal components: PC1 PC2
2KAX	Apo	S100A5_human	13-20 31-38 52-59 69-76	132.2 42.6 128.7 148.5 39.2 169.7	91.53 20.63
2KAY	Ca2	S100A5_human	13-20 31-38 52-59 69-76	135.2 103.4 130.7 119.7 30.1 121.1	25.09 -9.67
1ZFS	Ca2	S100A1_rat	14-21 30-37 54-61 72-79	140.7 98.5 129.8 120.0 35.7 119.8	30.27 -7.09
2RGI	Apo	S100A2_human	13-20 31-38 57-64 72-79	127.7 35.8 115.6 147.2 43.4 150.8	82.76 24.03
2Q91	Ca2	S100A4_human	13-20 31-38 55-62 72-79	137.6 119.7 127.7 101.6 32.7 107.5	7.10 -23.50
3C1V	Ca2	S100A4_human	13-20 31-38 55-62 72-79	137.9 118.8 127.3 102.0 34.0 108.5	8.35 -22.86

3CGA	Ca2	S100A4_human	13-20 31-38 55-62 72-79	135.4 119.0 130.4 102.8 34.6 106.9	6.38 -19.70
3PSR	Ca2	S100A7_human	12-19 28-35 54-61 71-78	135.1 102.6 120.3 121.1 33.9 130.2	29.60 -13.58
1YUR	Apo	S100A13_human	14-21 35-42 59-66 73-80	111.6 37.3 113.2 142.6 43.5 146.8	71.38 25.66
1YUT	Ca2	S100A13_human	17-24 35-42 56-63 73-80	139.1 107.4 127.5 113.4 27.7 118.2	22.36 -17.71
2EGD	Ca2	S100A13_human	17-24 35-42 56-63 73-80	135.9 113.1 129.4 109.9 30.9 112.2	13.68 -17.04
2H2K	Ca2	S100A13_human	17-24 35-42 56-63 73-80	135.0 112.6 128.7 111.1 31.3 113.9	14.45 -16.43
2PRU	Apo	S100B_human	9-16 29-36 56-63 70-77	115.8 61.9 108.1 152.8 45.4 160.6	63.83 17.29
1SYM	Apo	S100B_rat	10-17 29-36 56-63 70-77	130.7 16.1 119.7 145.2 32.6 135.0	89.76 29.34
3CR2	Ca2	S100B_bovine	11-18 29-36 53-60	136.2 125.7 125.2	-0.29 -28.07

			70-77	97.9 31.1 103.1	
2H61	Ca2	S100B_human	11-18 29-36 53-60 70-77	134.1 125.6 123.1 99.8 32.5 105.7	-0.01 -27.20

3.3

New insights into the mechanism of RAGE activation by S100 proteins (in preparation)

Introduction

The Receptor for Advanced Glycation Endproducts (RAGE) is a multiligand receptor that interacts with different extracellular ligands (AGEs, amphoterin, β-amiloyd peptide, S100 proteins)[1-5] and acts as a sensor for extracellular signals involved in the regulation of cell homeostasis and pathological states, including diabetic complications, tumour outgrowth, chronic inflammation and neurodegenerative disorders like Alzheimer's disease or multiple sclerosis[6-12]. RAGE is highly expressed during development[13]. In adult tissues its expression is very low. High expression levels of RAGE are associated with various tumors[14, 15], including colon adenocarcinoma and pancreatic carcinoma[16, 17], where RAGE is involved in tumour development and metastasis.

RAGE is an immunoglobulin-like cell surface receptor[18] with an extracellular region composed of three immunoglobulin domains (V – C1 – C2), a short transmembrane helix and a C- terminal cytosolic tail involved in signal transduction[19]. It was originally identified and characterized by its ability to bind advanced glycation end products (AGEs). Subsequently, a large number of RAGE ligands were identified, including S100 proteins, amphoterin/high mobility group box-1, amyloid-β, Mac-1[20]. However, the roles of individual ligands in the activation of RAGE are not fully understood.

This work focuses on studying the interactions between RAGE and S100 proteins. In particular we aim to find a key feature in S100 protein-RAGE interactions that can explain the differences reported for the RAGE signal transduction pathways induced by different members of S100 protein family despite the fact that different S100 proteins share a high degree of structural homology.

The S100 proteins are non-ubiquitous small acidic proteins (10–12 kDa) belonging to the EF-hand calcium-binding family, with 25–65% identity at the amino acid level and are found exclusively in vertebrates[21, 22], indicating that they are phylogenetically new proteins. In human genome, at least 25 members of the S100 proteins are known. Most of these genes (S100A1–S100A18, trichohylin, filaggrin and repetin) cluster to chromosome 1q21, [23, 24] known as the epidermal differentiation complex, which is frequently rearranged in human cancer[25]. Other S100 proteins are found at chromosome loci 4p16 (S100P), 5q14 (S100Z), 21q22 (S100B) and Xp22 (S100G) [26].

In vivo and *in vitro* experiments have shown that the S100 proteins can form non covalent homo- and hetero-dimers[27] and oligomers[28, 29]. Besides calcium, some members of the family also bind calcium zinc and/or copper[30, 31], which triggers structural changes

implicated in the interaction with target proteins and in modulation of the oligomaric state, contributing to their functional diversity.

S100 proteins regulate diverse cellular functions both in the intracellular and extracellular environments. They are implicated in the regulation of protein phosphorylation[32], enzyme activities[33], dynamics of cytoskeletal components[34], transcription factors[35], Ca^{2+} homeostasis[36], cell proliferation and differentiation[37]. Some S100 proteins are released into the extracellular space by an unknown mechanism. Extracellular S100s can act in a cytokine-like manner, as leukocyte chemoattractants, activators for macrophage and other inflammatory cells and modulators of cell proliferation[38]. These functions associate S100 proteins with a variety of pathologies such as inflammation, cardiomyopathies, and carcinogenesis. Therefore, during the last years, S100 proteins have attracted much attention as significant markers for diagnostic and clinical purposes and as therapeutic drug targets.

Many evidences report RAGE as the common receptor for extracellular S100 proteins. A large number of S100 proteins interact with RAGE extracellular domains *in vitro*. In cell-based assays some S100 proteins trigger RAGE-dependent signalling[39, 40, 41]. However, the molecular mechanism of the interaction is not clearly known, since several factors affect the interactions between RAGE and S100 proteins and the pathways triggered thereby. The RAGE binding properties of S100 proteins and the cellular effects of RAGE activation are modulated by different factors such as the oligomeric state of S100 proteins, their extracellular concentrations and metal binding. Moreover post-translational modifications such as glycosylation of the receptor or N-(carboxymethyl)lysine (CML) modifications of S100 proteins could also play important roles in defining specificity of interactions, oligomerization, and signalling. The S100-dependent RAGE activation and the effects on cellular activity have been investigated by several authors and summarized by Leclerc et al. and Salama et al. (Table 1). However, a standardized cell-based assay and systematic NMR analysis are needed to improve the understanding of the RAGE- S100 interactions.

Table 1: Summary of S100-protein dependent activation of the RAGE receptor

S100	Intracellular	Extracellular	RAGE domain	Technique	Cellular response	RAGE activation
S100B	Yes	Yes[1-4]	V[14] VC1[14] sRAGE[15,16] VC1[14,21]	SPR NMR	increases cell survival of SH-SY5Y cells at micromolar concentration[16]	mediated by V and C1 domains
S100A1	Yes	Yes[5]	GST-RAGE[17] V[18]	SPR	increased cell survival in a RAGE-dependent manner[22]	
S100A2	Yes	N.D.	GST-RAGE[17] V[18]	SPR	inhibitory influence on cell motility in squamous cell carcinoma lines[23]	
S100A3	Yes	N.D.				
S100A4	Yes	Yes[6]	GST-RAGE[17] sRAGE-Fc chimera[19]	SPR SPR	Stimulation of hPASMCs resulted in an increase in migration blocked by anti-RAGE[24]	
S100A5	Yes	N.D.	V[18]	SPR		
S100A6	Yes	Yes[7]	V[20] VC1[20] C2[20] sRAGE[20]	SPR	mediated apoptosis of the SH-SY5Y cells at micromolar concentration[20]	mediated by C1 and C2 domains
S100A7	Yes	Yes[8]	N.D.[25]	In vitro ligand - receptor intreaction assay[25]		
S100A8/A9	Yes	Yes[9]	RAGE is involved in cell growth signaling by S100A8/A9[26]	RAGE knock-down experiments; protease protection assay	S100A8/A9 at low micromolar concentrations promotes cell growth in different tumor cells[26]	
S100A10	Yes	N.D.				
S100A11	Yes	Yes[10]	N.D.[10,27]	ko mice; cell assays		
S100A12	Yes	Yes[11]	sRAGE-Fc[19] V[18] C1C2[28] sRAGE[28] C1C2[28] RAGE-Fc fusion protein[29]	SPR TNS fluorescence native PAGE NMR SPR		
S100A14	Yes	N.D.				
S100A13	Yes	Suggested[12]	N.D.[30]	cell assay		
S100A14	Yes	N.D.				
S100A16	Yes	N.D.				
S100P	Yes	Yes[13]	N.D.[31-33]	co-immunoprecipitation; cell assays	increased cell proliferation in a colon cancer cell line[33] stimulates proliferation, survival, motility, and invasiveness of a pancreatic cell line and associated with high levels of tumor growth and metastasis in vivo[31] increased cell proliferation and survival on NIH3T3 cells[32]	
S100Z	Yes	N.D.				
S100G	Yes	N.D.				

In this work the RAGE signal transduction pathways induced by different members of S100 proteins family has been investigated in vivo and in vitro. A cell-based assay was performed to investigate the interaction between seven recombinant S100 proteins and RAGE in two

models of human cancer cell lines (SW480 and PANC-1) that express RAGE. The RAGE-dependent effects of different S100 proteins and the role of extracellular RAGE domains on tumour cell proliferation were compared. To further improve the understanding of the structural basis of RAGE activation by S100 proteins, the interaction was investigated by NMR and a structural model for thye complex of S100P with V domain of RAGE has been calculated using HADDOCK program.

Materials and methods

Cell Culture

The colorectal adenocarcinoma cell line SW480 and the pancreas carcinoma cell line PANC-1 were obtained from the American Type Culture Collection (ATCC). Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (PAA) supplemented with 10% foetal bovine serum (FBS) (PAA), 2 mM glutamine (Lonza), penicillin (100U/ml) and streptomycin (100 U/ml) (Lonza), and were maintained in a humidified incubator containing 5%CO_2 at 37°C.

Expression and Purification of Recombinant S100 proteins and RAGE domains

S100 A1, A2, A5, A12 were cloned in pET 21a using the restriction enzymes *Nde*1 *and Xho*1 in the native form. S100A4 was cloned in pET DEST42. A stop codon was introduced at the 3' ends of the gene sequences to eliminate the histidine tag of the vector. S100P was cloned in pETG-30A to obtain a fusion protein with N-terminal histidine-GST tag. The recombinant plasmids with S100 genes were transformed into *Escherichia coli* Bl21 (DE3) Gold cells except for S100A4 which was expressed in pLys S cells. For protein expression, cells were grown at 37°C to an optical density of 0.6 at 600 nm and expression was induced with IPTG (0.5-1 mM). After induction, the growth temperature was reduced to 25°C in case of S100A4 and S100A5, S100P. Cells were harvested 12-16 hours after induction in all cases and resuspended in lysis buffer- 20 mM Tris, 2 mM DTT, 1 mM PMSF, pH 7.5. Lysis was performed by sonication. The suspension was ultracentrifuged at 40000 rpm for 30 minutes. The supernatant was brought to 5 mM $CaCl_2$ and applied to a phenyl sepharose (high sub) column equilibrated with 20 mM Tris, 5 mM $CaCl_2$. Loading was done very slowly at a rate of 0.4-0.5 ml of lysate/ minute. The unbound proteins were washed out with 12-15 column volumes of equilibration buffer. S100 proteins were then eluted with 20 mM Tris, 10 mM EDTA, pH 7.5. The elute was concentrated and applied to a Superdex 75 16/60

column equilibrated with 30 mM MES, 100 mM NaCl, 5 mM DTT (in case of the S100s with cysteines), pH 6.5. The purest fractions were pooled and washed with EDTA to obtain the protein in apo form. His-tagged S100P was purified by a nickel sepharose column and the N-terminal tag was cleaved by TEV enzyme. The non-fused protein was purified by passing the mixture over a nickel column.

DNA coding for human soluble RAGE V domain (aminoacids 23-132), VC1 domains (amino acids 23-243) and C1C2 domains (aminoacids 122-347) was amplified via PCR with oligonucleotides containing 5′ NdeI and 3′ XhoI restriction sites. The amplified DNA sequences were cloned into the NdeI and XhoI sites of pET15b vector to express the proteins with an N-terminal hexa-histidine tag followed by a thrombin cleavage site. Recombinant plasmids with V, VC1, C1C2 RAGE domains were transformed in E. coli Origami(B) DE3 (Novagen). Cells were grown at 37°C till an O.D.of 0.7. Protein expression was induced with 0.5mM IPTG and growth was allowed for 6 hours at 20°C. Cells were harvested and lysed in 20mM phosphate, 500mM NaCl, 1mM PMSF, pH 7.4. The lysate was first purified on HiTrap chelating HP column (GE Healthcare) and eluted with 20 mM phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.4. Following dialysis in 20 mM phosphate pH 6.0, 150 mM NaCl, the His6 tag was removed by thrombin cleavage for an overnight period at room temperature. The protein without fusion tag was then purified by cation exchange in MonoS (GE Healthcare) with a linear gradient from 150-850 mM NaCl. Fractions containing the pure protein were pooled and concentrated. The C1C2 domain was expressed in E. coli BL21 (DE3) Gold cells in the insoluble fraction. The protein was refolded in the nickel column during purification. For cell line experiments, all proteins were prepared in PBS buffer, pH 7.4. For NMR experiments the same were prepared in 20 mM Hepes, 100 mM NaCl, pH 7.2. For titration of S100 proteins with more than one cysteine residues with RAGE, the S100s were first reduced by DTT which was then washed out under reducing conditions.

S100 proteins treatments

Cells were seeded in 96-well plates ($1x10^3$ cells/well). After 24 hours they were treated with various concentrations of recombinant S100 proteins (0.1 – 1 – 10 – 100 – 1000 nM) for 72 hours. For each S100 protein, the concentration with maximum effect on cell proliferation was determined, assayed by the MTS assay (Promega, Madison, WI, USA). In brief, 1000 cells in 100 µl medium were seeded in each well of 96-well plates. At the end of the 72-h incubation period with the recombinant S100 proteins, the cells were incubated with 20 µl of

a MTS/PMS mixture (MTS/PMS ratio: 20:1) in each well of 96-well plates for 2 h at 37 °C in a humidified incubator with 5% CO2 to allow viable cells to convert the tetrazolium salt into formazan. The conversion to formazan was determined by measuring the absorbance at 490 nm using a PerkinElmer Victor plate reader (PerkinElmer, Shelton, CT, USA).
The cells were treated with S100 proteins at the optimal concentration and their growth curve was evaluated between 24 and 120 h of treatment, by the MTS assay (Promega, Madison, WI, USA).

RAGE domains treatments

Cells were treated with S100 proteins along with different concentrations of recombinant RAGE extracellular domains (V, VC1, C1C2; 0.01-0.1-1-10-100 nM) and the optimal concentration of RAGE domains was established by the MTS assay. The RAGE domain which binds to each S100 protein and inhibits it from stimulating cell proliferation was determined.

NMR experiments

The experiments for protein assignment and mobility measurements of holo S100P and V-domain of RAGE where performed on protein samples at concentrations ranging between 0.4 and 1 mM on Hepes buffer at pH 7.2. NMR experiments were acquired on Bruker AVANCE 900, AVANCE 800, AVANCE 700 and DRX 500 spectrometers, equipped with triple resonance cryo-probes at 298 K and 310 K. All spectra were processed with the Bruker TOPSPIN software packages and analyzed by the program CARA (Computer Aided Resonance Assignment, ETH Zürich). The backbone resonance assignment of ^{15}N, ^{13}C enriched holo-S100P domain was obtained by the analysis of HNCA, HNCACB and CBCA(CO)NH spectra performed at 310 K on a 500 MHz spectrometer. The assignment of ^{1}H-^{15}N HSQC and ^{1}H-^{15}N TROSY S100P spectra at 298 K were derived from data at 310 K, integrated by the analysis of HNCA spectra at 298 K. The interaction of S100 proteins with V-domain, VC1- and C1C2-tandem domains was monitored by ^{1}H-^{15}N HSQC and ^{1}H-^{15}N TROSY spectra.

Results

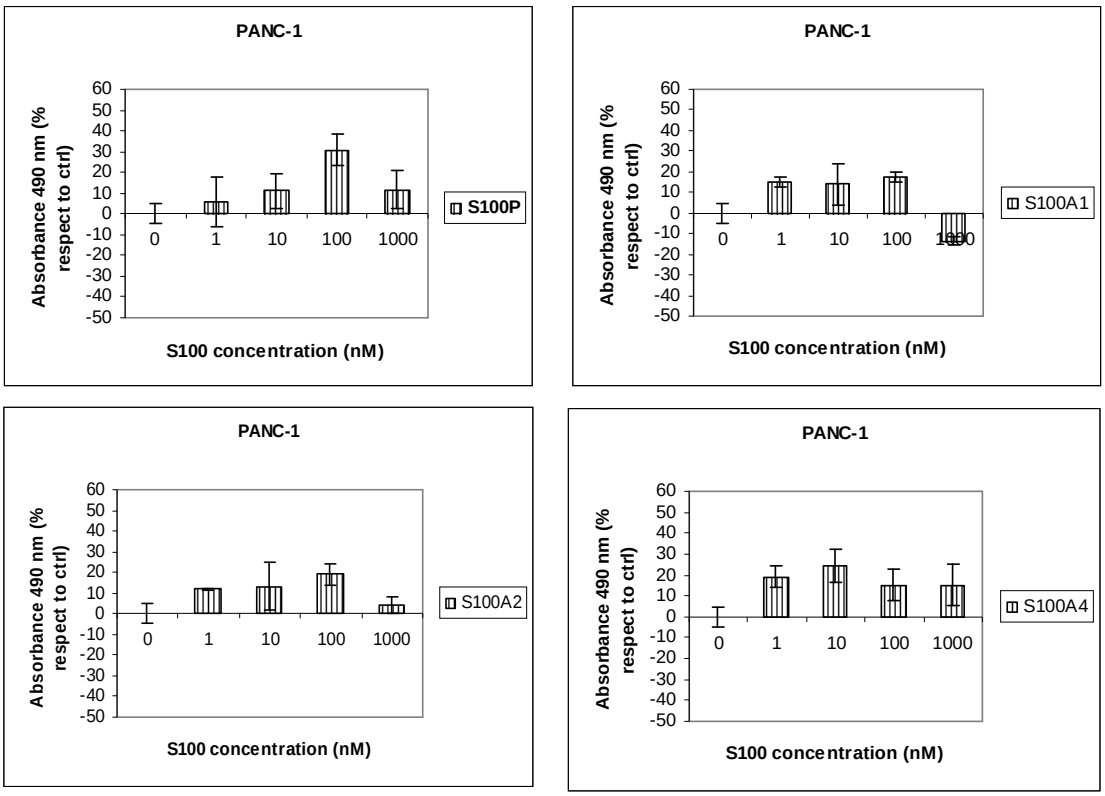
The SW480 and PANC-1 cell lines express RAGE on their membrane and previous studies have shown that the activation of this receptor by S100P stimulates cell proliferation in these cell lines. In our study we performed a cell-based assay comparing the effects of different

S100 proteins on the proliferation of these two well characterized tumour cell lines. Cells were seeded in a 96 well plate at a density of 1000 cells/well in 100 µl of medium with 10%

FBS. After 24 hours, the medium was removed and cells were treated with S100 proteins diluted in the serum-free medium.

The effect of recombinant S100A1, A2, A4, A5, A12, A16 and P on cell proliferation was assayed after 72 hours of treatment. For each cell line, a dose-response curve was plotted, treating the cells with the S100 proteins at concentrations ranging between 1 nM and 1 µM.

In both the cell lines, the S100 proteins triggered a dose-dependent effect on cell proliferation, except for S100A16, which stimulates the proliferation of SW480 but not of PANC-1 cell line. As shown in fig. 1, our data confirm the already published results for S100P, which exhibits maximum effect on cell proliferation at a concentration of 100 nM. Moreover, we found that some S100 proteins exert maximal effect at the same concentration (S100A1, A2, A12, A16), while others (S100A4, S100A5) are more effective at lower concentrations (10 nM).



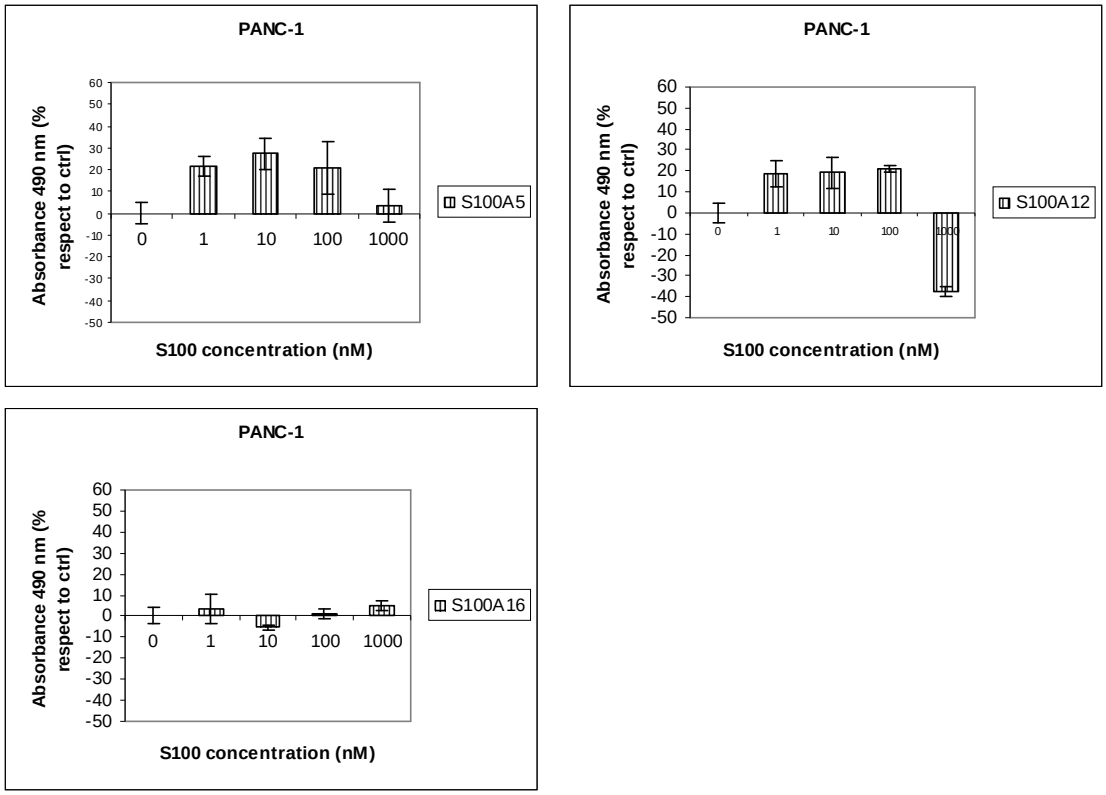
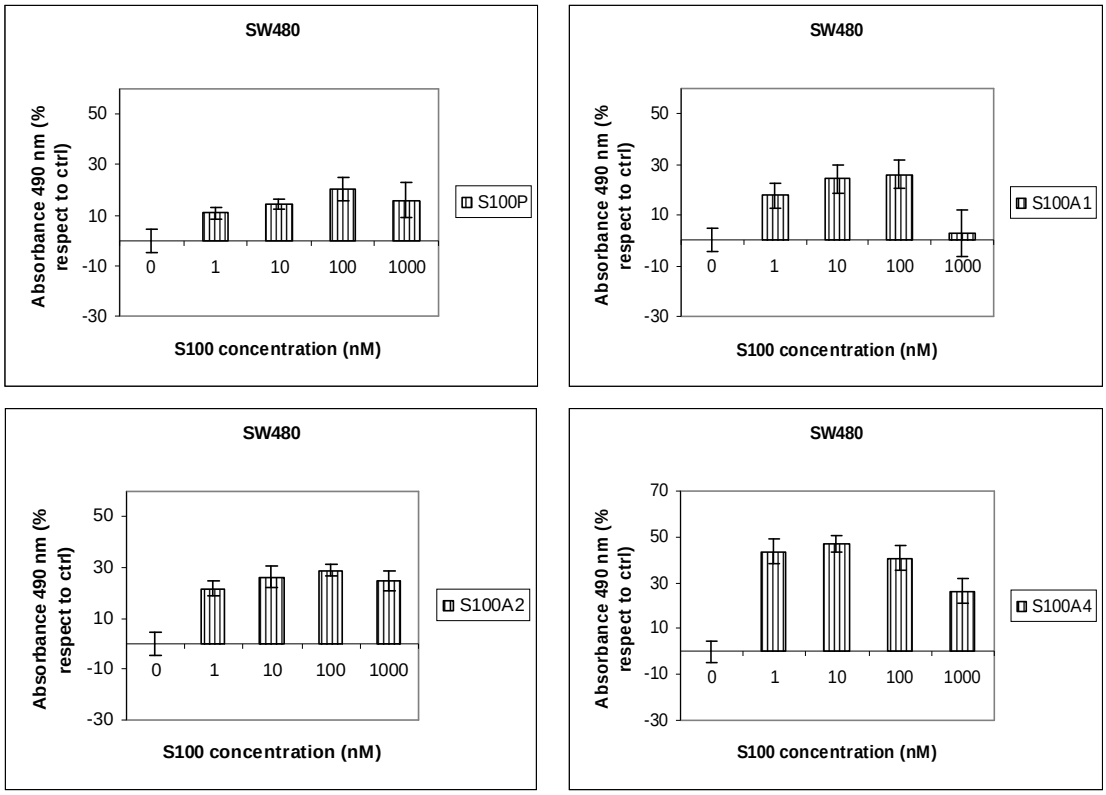


Figure 1A: Dose response curve of Panc-1 cells treated with different concentrations (0, 1, 10, 100, 1000 nM) of S100 proteins.



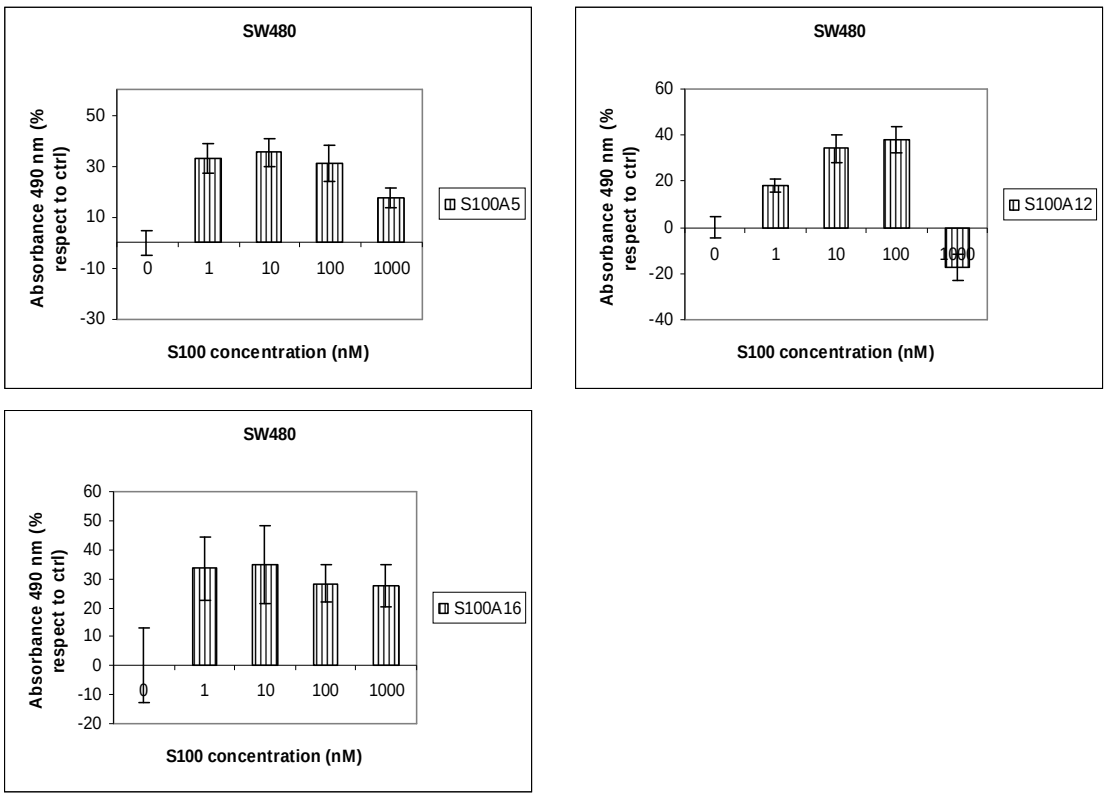


Figure 1B: Dose response curve of SW480 cells treated with different concentrations (0, 1, 10, 100, 1000 nM) of S100 proteins.

From our data we can also elicit that S100A12 at micromolar concentration has an inhibitory effect on cell proliferation, as already reported for S100B, while the other S100 proteins have a positive effect on cell proliferation at any tested concentration.

The effect of S100 proteins on cell proliferation was also evaluated by measuring the growth curve of SW480 and PANC-1 treated for 24-96 hours with a single concentration of S100 proteins.

The cells were seeded as reported previously and treated with 100 nM S100P, A1, A2, A12 and A16; with 10 nM S100A4 and A5 and with 1 μM S100A12. Results shown in fig. 2 confirm the stimulatory effect of nanomolar concentrations of S100 proteins in both SW480 and PANC-1 cell proliferation, as observed for the dose-response curves measured at 72 hours. S100A12 has different effects on cells proliferation at different concentrations. The cells treated with 100 nM S100A12 grow faster than untreated cells while cells treated with 1 μM S100A12 have a lower growth rate respect to untreated cells.

For S100A16 a stimulatory effect is observed on SW480 cell proliferation but not on PANC-1.

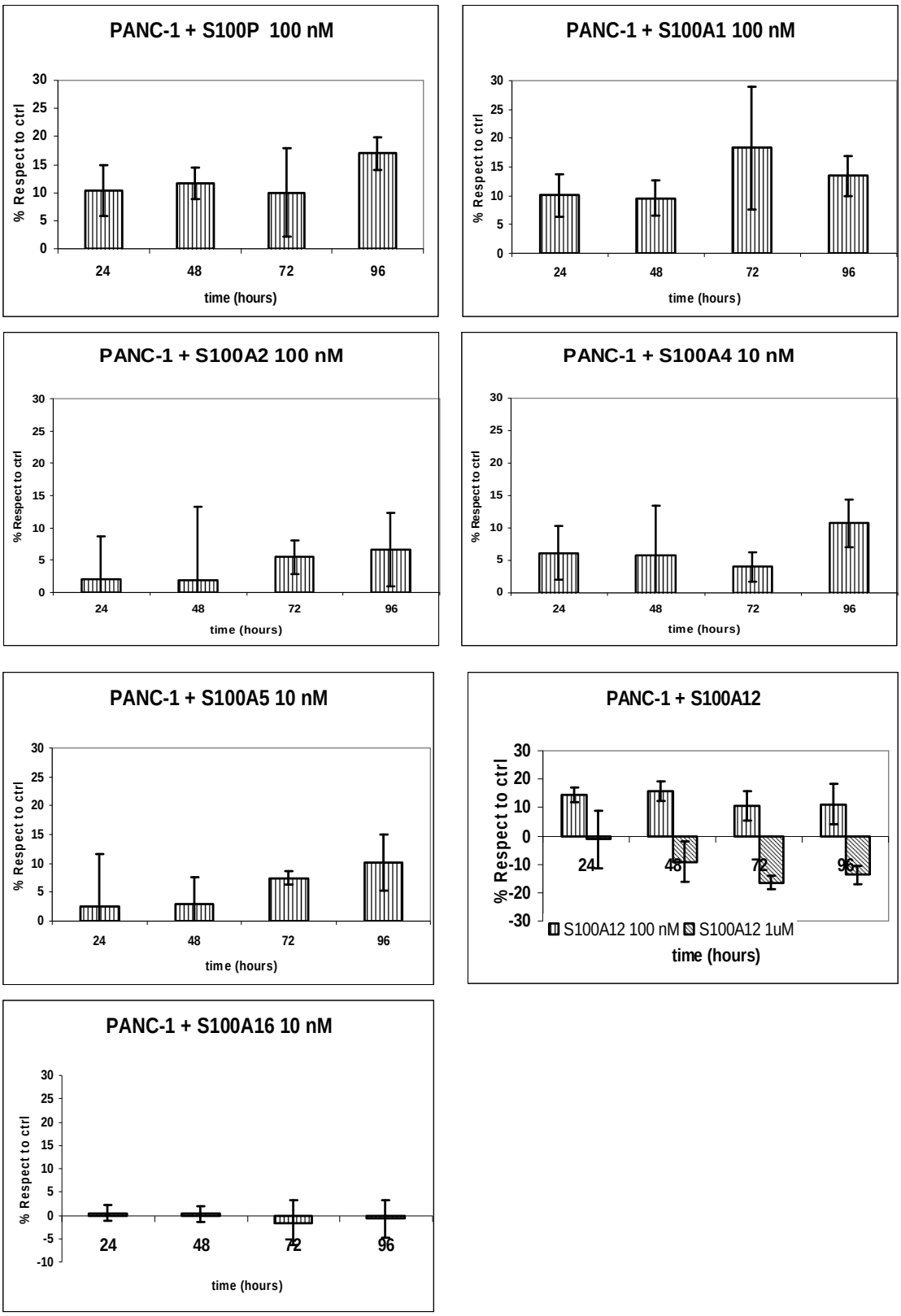


Figure 2A: Growth curves of Panc-1 cells treated with different S100 proteins upto 96 hours.

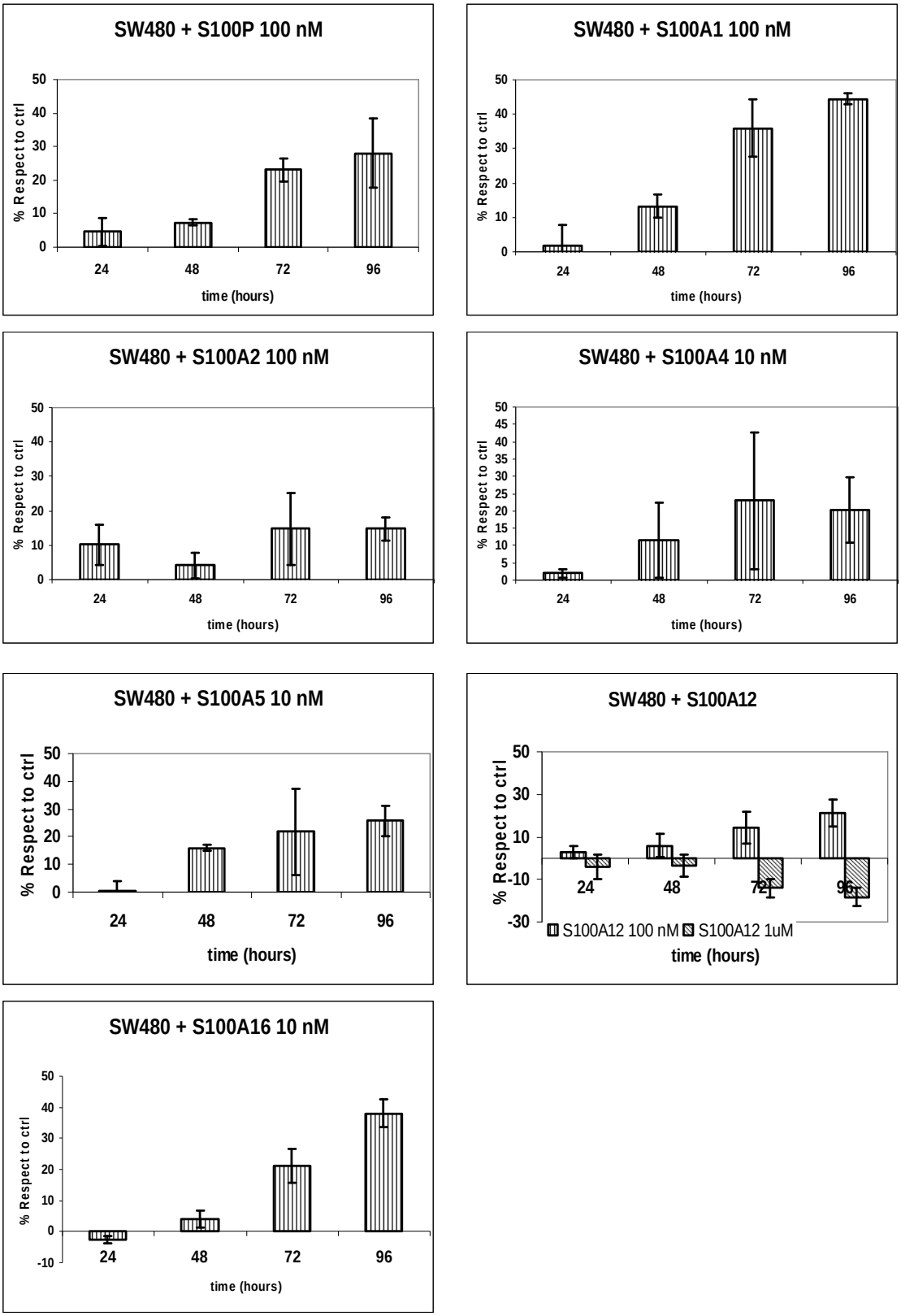
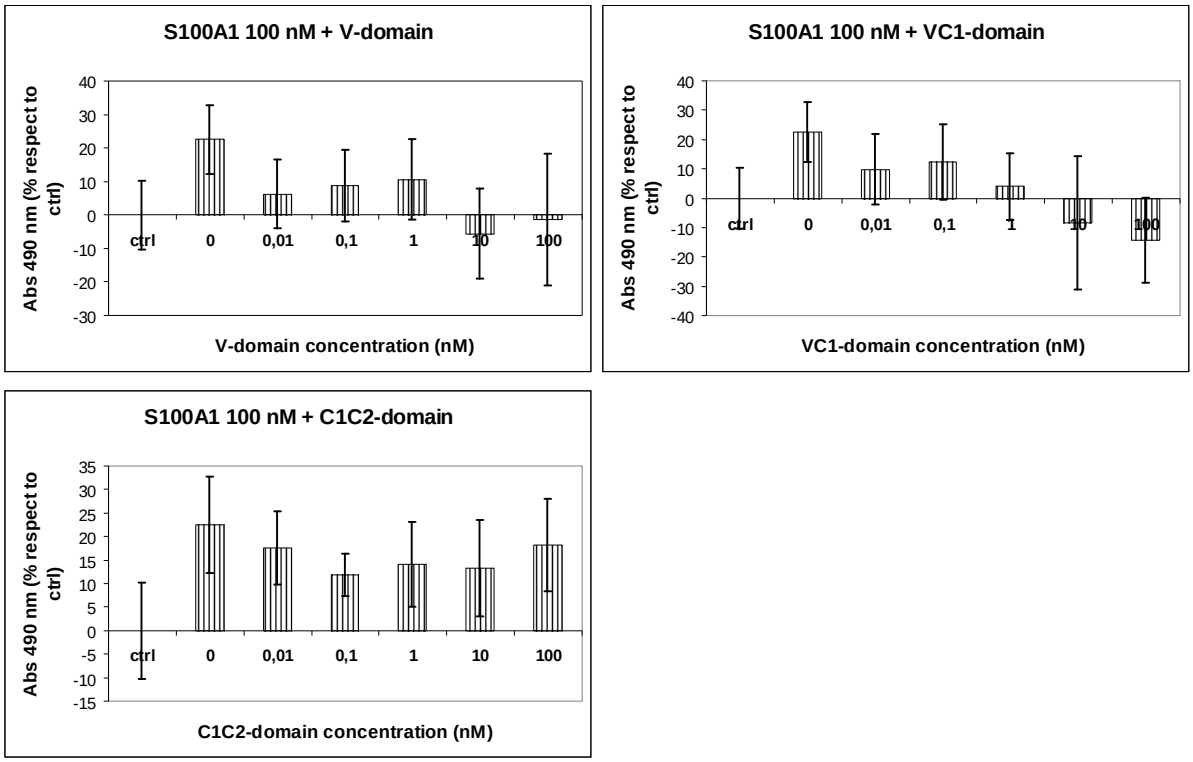
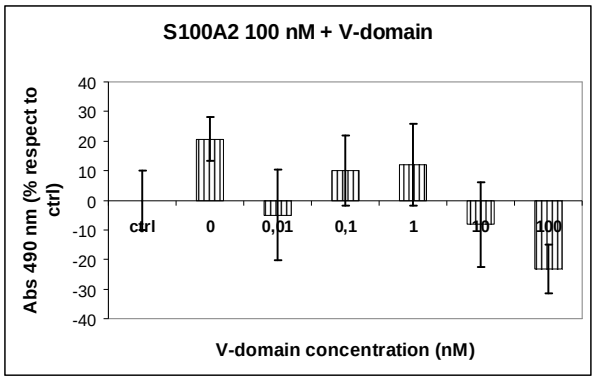


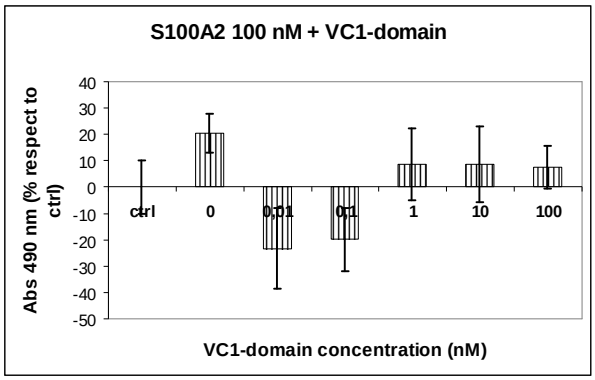
Figure 2B: Growth curves of SW480 cells treated with different S100 proteins upto 96 hours.

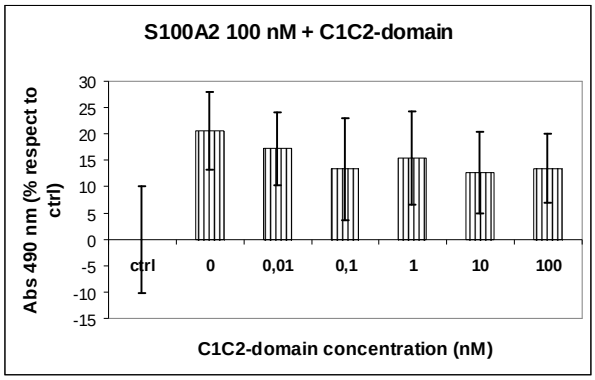
Next, the role of the different domains of RAGE in ligand binding was investigated through competition experiments on the same cell lines and NMR studies in vitro. Cells were treated with both S100 proteins and soluble RAGE domains (V, VC1, C1C2) and the competition between soluble and membrane RAGE was expected to modulate the effect of S100 proteins on cell proliferation. S100A1, A2, A5, P are reported to interact with RAGE V-domain, as assayed by SPR experiments. Our competition experiments (fig. 3) confirmed these interactions, since the presence of V or VC1 domains but not C1C2 inhibit cell proliferation triggered by S100A1, A2, A5 and S100P. On cells treated with 100 nM S100A12 only VC1 and C1C2 but not V domain inhibits the stimulatory effect on cell proliferation, providing new evidences on the pivotal role of C1 domain in the binding of this S100 protein.



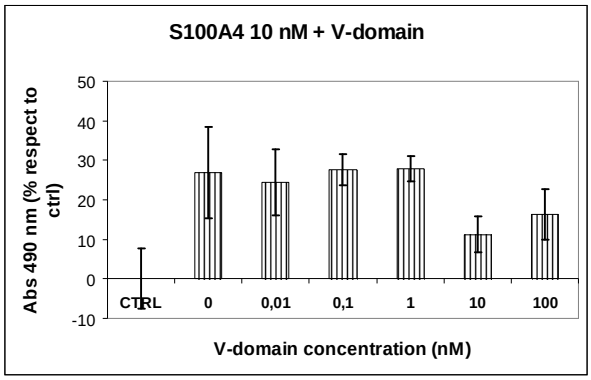
V and VC1 domains of RAGE inhibit S100A1-stimulated cell proliferation but not C1C2. So S100A1 interacts with the V-domain of RAGE.

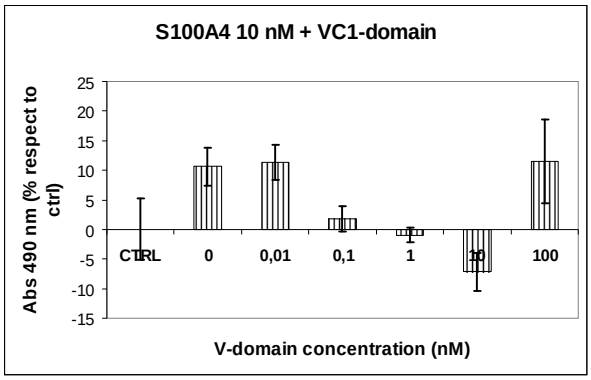


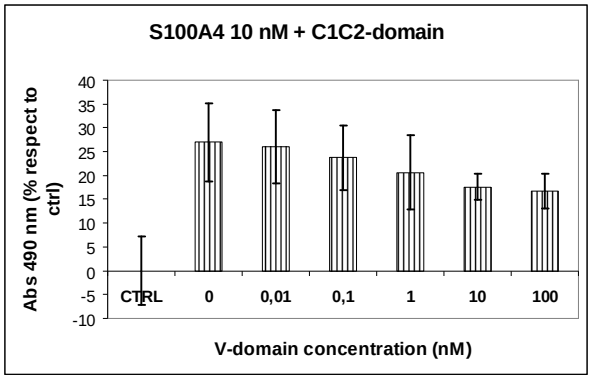




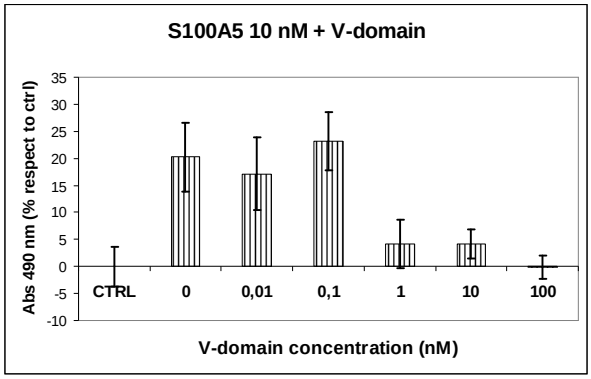
V and VC1 domains of RAGE inhibit S100A2-stimulated cell proliferation but not C1C2. So S100A2 interacts with the V-domain of RAGE.

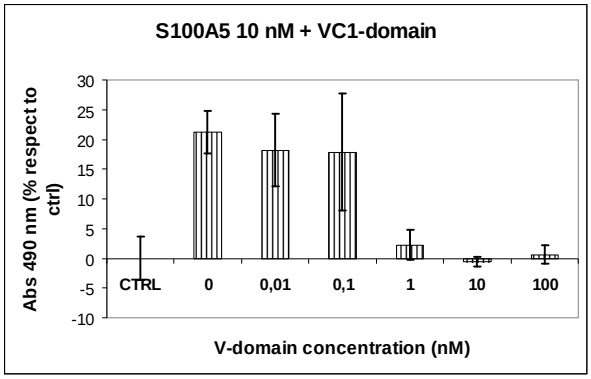


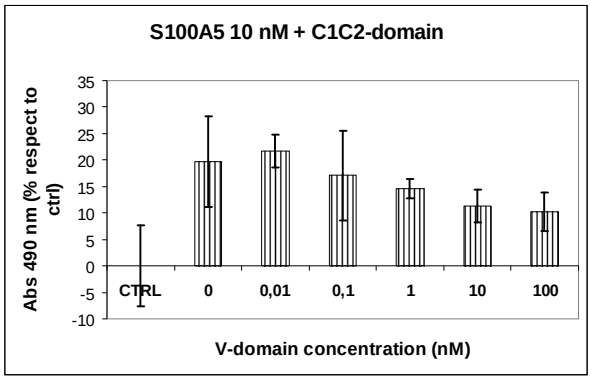




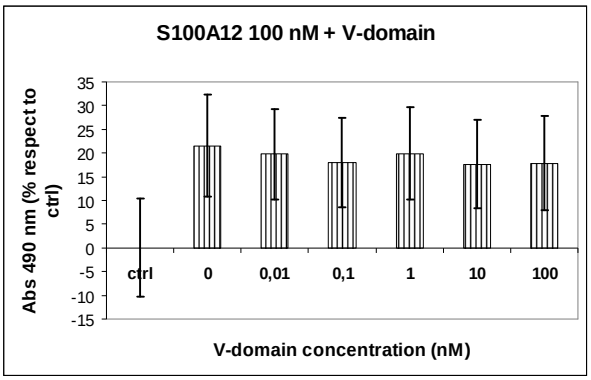
V and VC1 domains of RAGE inhibit S100A4-stimulated cell proliferation but not C1C2. With V-domain the inhibitory effect is less. So S100A2 interacts with the V-domain of RAGE with low affinity and with VC1 tandem domain with a higher affinity.

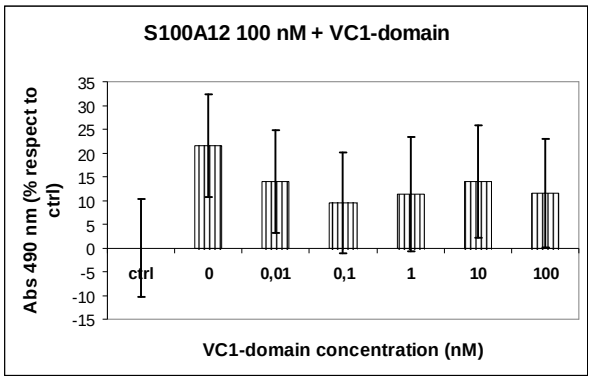


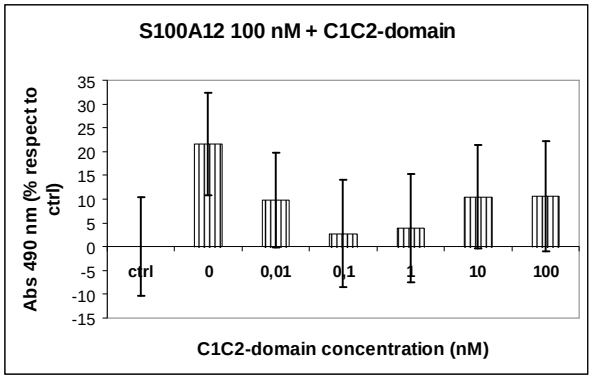




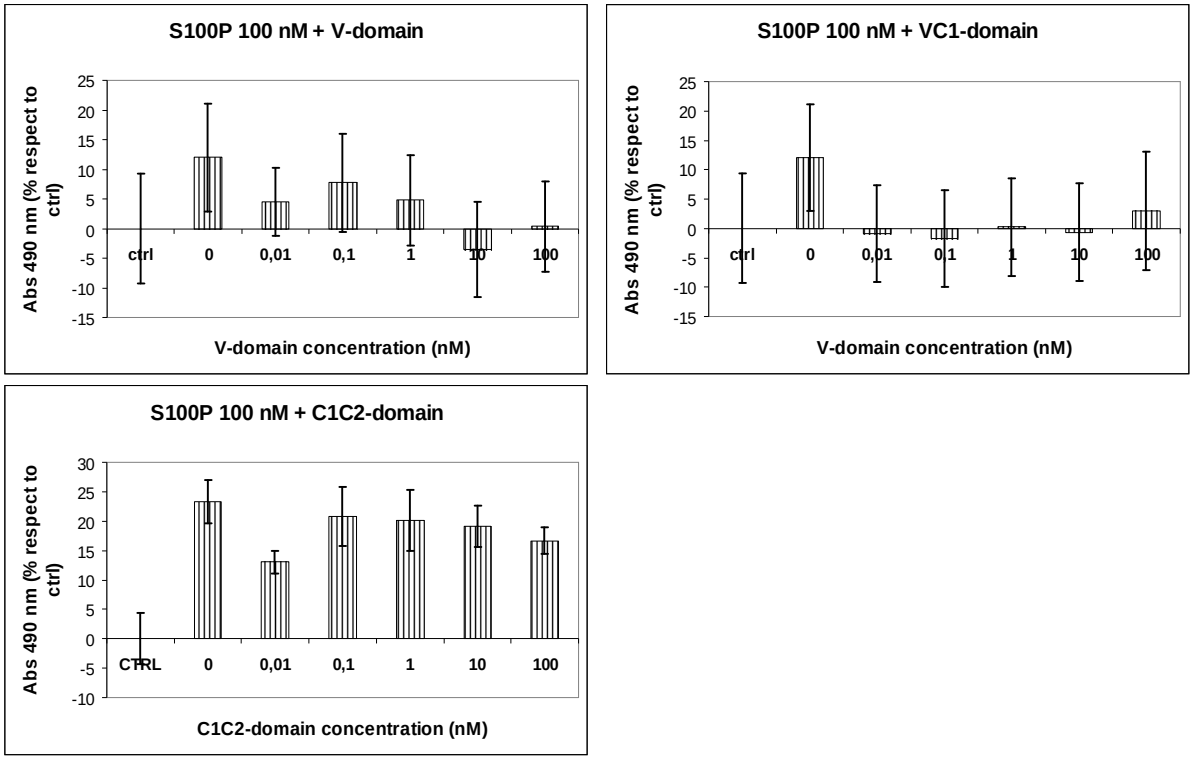
V and VC1 domains of RAGE inhibit S100A5-stimulated cell proliferation but not C1C2. So S100A5 interacts with the V-domain of RAGE.







C1C2 domain of RAGE inhibits S100A12-stimulated cell proliferation but not V or VC1 domains. So S100A12 interacts with the C1C2-domain of RAGE.



V and VC1 domains of RAGE inhibit S100P-stimulated cell proliferation but not C1C2. So S100P interacts with the V-domain of RAGE.

Figure 3: Growth curves of Panc-1 cells from competition experiments: effect of soluble RAGE domains on cells treated with S100 proteins.

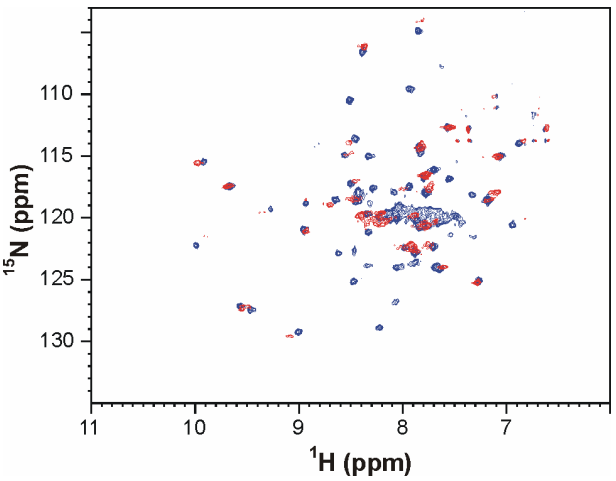
Interesting are also the results collected on S100A4. The interaction of this protein with RAGE was investigated here for the first time. Our data show that the addition of VC1 inhibits the effect of S100A4 on cells while V-domain is effective only at higher concentration and C1C2 doesn't have any effect. All together these data suggest that S100A4 interact with VC1 tandem domain and with V domain with lower affinity but not with C1C2 tandem domain.

Table 2: Effect of S100 proteins on cell proliferation and RAGE activation

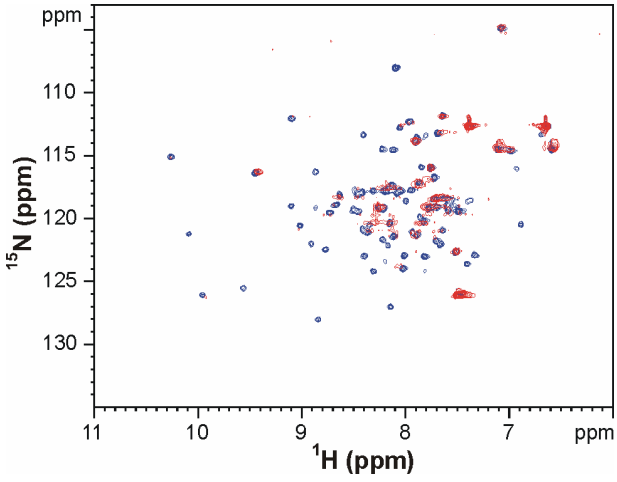
S100 protein	Increased PANC-1 proliferation	Increased SW480 proliferation	Competitive RAGE domain	NMR interaction
S100A1	yes	yes	V - VC1	
S100A2	yes	yes	(V)[b]- VC1	V
S100A4	yes	yes	V - VC1	V
S100A5	yes	yes	V - VC1	V-VC1
S100A12	yes/no[a]	yes/no[a]	(VC1)[b] - C1C2	C1C2
S100A14	no	no		
S100A16	no	yes		V
S100P	yes	yes	V - VC1 - C1C2	V-VC1

The real binding capability of the investigated S100 proteins toward RAGE domains was then assessed in vitro by NMR.

Alteration of chemical shift, signal disappearance and line broadening in ^{1}H-^{15}N HSQC spectra are considered as evidences of interactions between S100 proteins and RAGE. ^{15}N labelled samples of Ca^{2+}- S100A2, -S100A5, -S100A12 and -S100P were titrated with unlabeled V-domain of RAGE. All the investigated proteins but S100A12 exhibit a general broadening of resonances with disappearance of several peaks and several shifts when V-domain was added to a molar ratio of 1:2 (V-RAGE:S100). Similar effects were observed when C1-C2 tandem domain of Rage was added to Holo-S100A12.



Ca^{2+}-S100A2 Ca^{2+}-S100A2 + V-RAGE



Ca^{2+}-S100A5 Ca^{2+}-S100A5 + V-RAGE

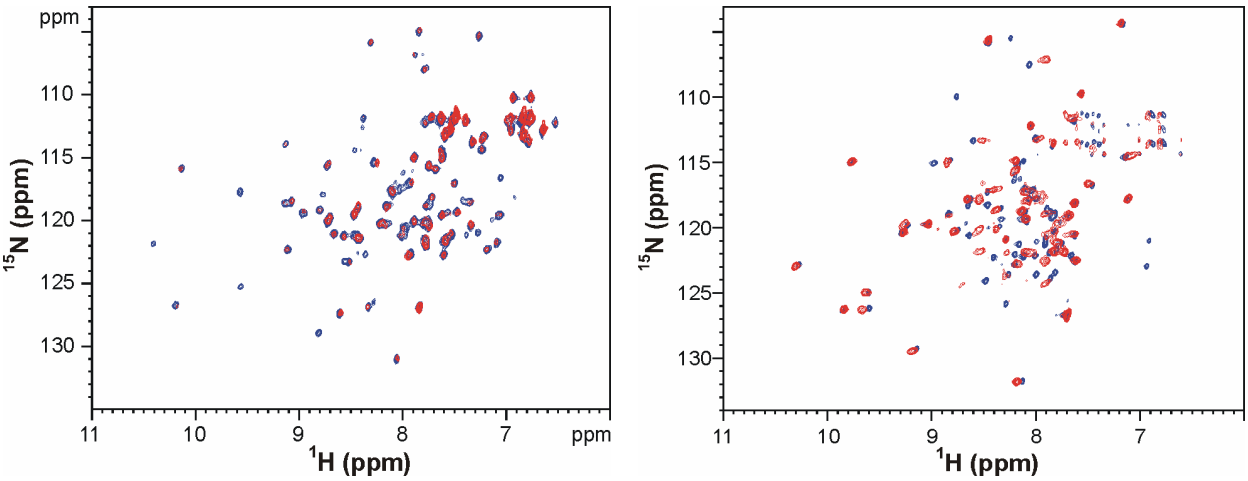


Figure 4: Overlay of 1H ^{15}N HSQC spectra of Ca2^{+}-S100 proteins titrated with RAGE domains in 2:1 ratio. Buffer conditions: 20 mM Hepes, 100 mM NaCl, pH 7.2. Temp:298 K. Spectra taken at 800 MHz.

The interaction of holo-S100P with extracellular domains of Rage was investigated in more detail.

Structural data representative of interactions between S100P and V-domain of RAGE in solution were obtained by monitoring the reciprocal effects on ^{1}H-^{15}N HSQC and Trosy NMR spectra. The backbone assignment of apo S100P was already available on BMRB database while the holo form was assigned in the present work. Backbone resonances of V-domain of RAGE were also reassigned. The spectral quality for holo S100P is largely better at 310 K than at 298 K. Thus all the experiments for backbone assignment were performed at the physiological value of 310 K. Holo S100P exists as homodimer in solution and during size exclusion is eluted as a single peak.

Measurements of longitudinal (R_1) and transverse (R_2) relaxation rates of backbone amide nitrogens at 700 MHz ^{1}H Larmor frequency and 310 K have been performed on a ^{15}N-enriched sample of holo S100P. Estimates of R_1 and R_2 values under the chosen experimental conditions of magnetic field and temperature were obtained using the program HydroNMR (48) and the X-ray structure taken from PDB (30).

The experimental and theoretical R_1 and R_2 values are reported in Figure 5. The observed R_1 and R_2 data are in excellent agreement with the theoretical expectations for a dimeric form of S100P without any indication of self aggregation at this concentration.

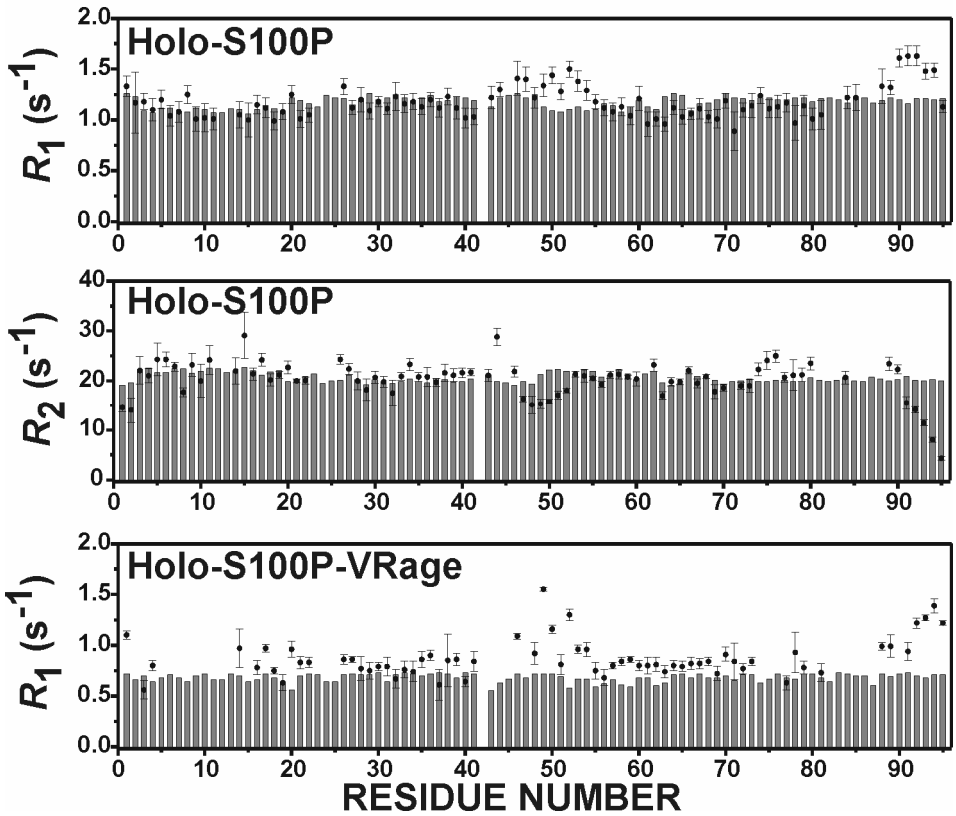


Figure 5: NMR relaxation data for Holo-S100P and for its adduct with V-RAGE (data collected at 310 K). The calculated (grey bars) and experimental (filled circles) backbone ^{15}NH R_1 and R_2 values for the isolated Holo-S100P are shown in A and B, and those of its complex with V-RAGE in C. The agreement between experimental and calculated R_1 and R2 values for the Holo-S100P is excellent and only few residues at the C-terminal and some residues on the loop between the two EF-hand motif experience fast internal mobility revealed by the large R_1 and the small R_2 values. Also the experimental R_1 data for the Holo-S100P/V-RAGE adduct agrees with the theoretical expectations while R_2 are not reported because the large errors due to the low signal-to-noise ratios in the spectra.

The titration of ^{15}N labeled holo-S100P with unlabeled V-domain until a molar ratio of 2:1, resulted in a progressive line broadening of peaks, with shifts and disappearance of signals. No further changes in the NMR spectrum of holo-S100P were detected at higher concentrations of V-domain. To improve the quality of the spectra the titration was repeated on a ^{2}H-^{13}C-^{15}N sample of holo-S100P.

Comparable effects on holo-S100P spectra were obtained by performing the titration with VC1-tandem domain. To complete the analysis of the protein-protein interface, a ^{15}N labeled sample of V-domain was titrated with holo-S100P until a molar ratio of 1:2. The analysis of the corresponding ^{15}N HSQC spectra revealed a set of signals experiencing chemical shift variation. Chemical shift changes on holo-S100P and V-domain are reported in figure 6.

From the analysis of the ^{1}H and ^{15}N chemical shift variations between isolated holo-S100P and the complex it is clearly evident that the protein regions formed by Helix-4, Helix-3 and the loop interconnecting Helix-3 with Helix-2 are strongly affected by the interaction.

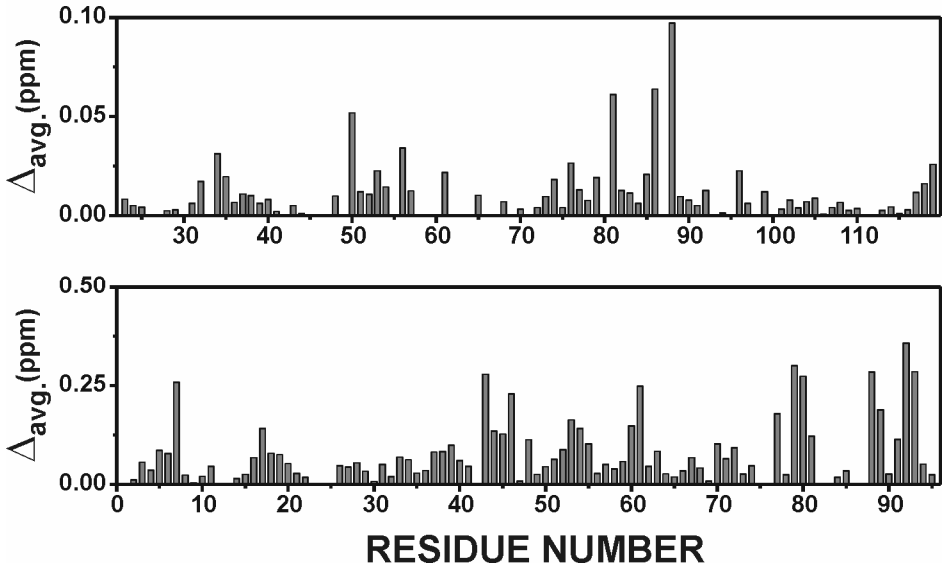


Figure 6: Plot of the ^{1}H-^{15}N HSQC chemical shifts ($\{[(\delta_H)^2 + (\delta_N/5)^2]/2\}^{1/2}$) per residue observed in 0.4 mM Holo-S100P in presence of V-RAGE (0.2 mM) (**A**), and in 0.2 mM of V-RAGE in presence of Holo-S100P (0.4 mM) (**B**).

The chemical shift mapping and the related interaction surfaces have been exploited to determine a possible structural model for the interaction of the holo-S100P with V-domain of RAGE using the program HADDOCK (Dominguez et al., 2003). To model the structure of the complex between holo-S100P and the V-domain of RAGE, we performed docking calculations using as input a computer homology model of holo-S100P based on the X-ray structure 1J55 and the X-ray structure of the VC1 tandem domains. NMR data were used to derive the interaction restraints on the two proteins. From the analysis of the energetic and scoring functions, it appeared that, essentially, only one cluster experienced a far better scoring function with respect to the others with larger interaction surfaces and a much lower average number of ambiguous restraints. This indicates better agreement with the experimental data. The lowest-energy complex for this cluster is shown in Fig. 7.

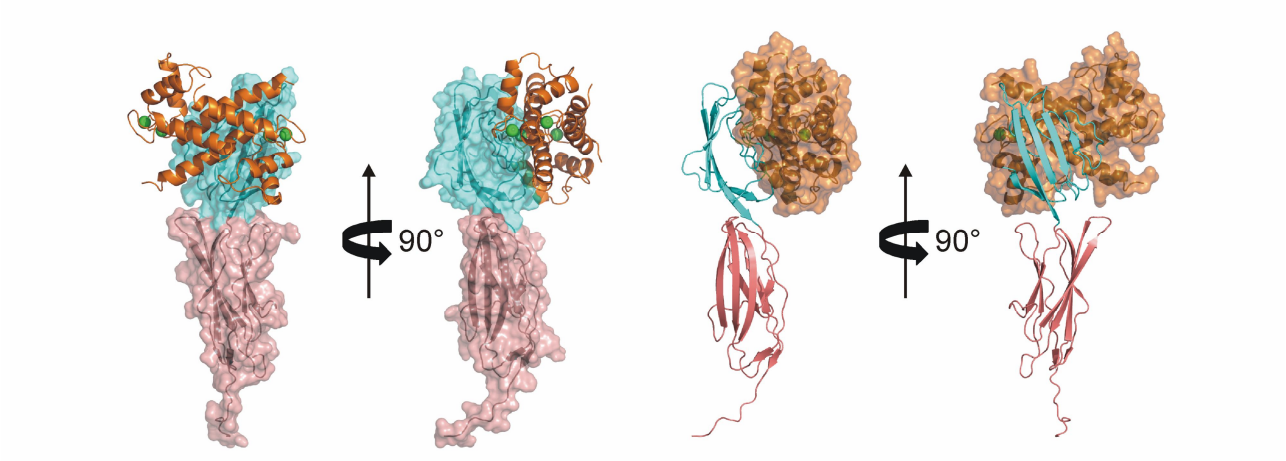


Figure 7: Structural models of Holo-S100P-VC1 complex generated by molecular docking. Lowest HADDOCK score structure of cluster 1 is represented. Holo-S100P is shown in orange, while VC1 domains are shown in cyan (V) and pink (C1).

Discussion

Some members of S100 family such as S100B, S100P and S100A12 are secreted or released in the extracellular space where they activate the RAGE signaling pathway by interacting with the extracellular domains of the receptor. The interaction with RAGE is supposed to be the general mechanism for the autocrine and paracrine activity of S100 proteins. However, the experimental data on activation of RAGE by these calcium binding proteins available so far are too scattered and still largely incomplete. Our results on a broad set of S100 show that all the investigated proteins trigger the RAGE signaling increasing cell proliferation with different potency. S100A4 and S100A5 exhibit the maximum activity already at concentration of 10 nM while all the others require ten times larger concentration. The NMR data in vitro and the competition experiments carried out on the same cell lines support a model whereby all the S100 proteins induce an increase of cell proliferation by a direct protein-protein interaction with the extracellular domains of RAGE. The NMR spectra show that all the investigated holo-S100s interact with at least one of the extracellular domains of RAGE with an affinity in the micromolar range or lower. In particular, all but S100A12 exhibit relevant alteration of the 1H-15N HSQC spectra in presence of V or VC1 domains. The role of V, VC1 and C1C2 has also been assessed by performing competition experiments on cell culture where they abolish the proliferative effects of the S100 proteins to which they bind. From our results it is clear that V-domain plays the leading role in binding to most S100 proteins and in activation of the receptor. This can be explained by the homology of V domain with the region of immunoglobulin responsible for binding to

epitope of antigens and its structural features similar to cell adhesion molecules. Activation of RAGE involves oligomerisation of the receptor modulated by the ligand binding. According to the most recent hypothesis, pre-assembled RAGE receptors are tethered by S100 oligomers and switched in the activated form. In this respect, structural details of the binding mode of S100 ligands on the V-domain can provide information on the self-assembled S100 species involved in RAGE oligomerisation and on the mechanisms that tune extracellular activity of these calcium binding proteins.

The model here proposed for adduct between S100P and V-domain of Rage shows that the negatively charged surface of the calcium-binding protein in between helix 3 and 4, interact with the region of V-domain formed by strands 2, 5, 6 and by some interconnecting loops.

A similar binding mode has recently been proposed for S100B that interacts with the same negatively charged region on a positively charged area of V-domain not far from that here reported for S100P. Although without any functional relevance, it is interesting to note that in the available octameric S100B crystal structure the interacting surfaces responsible for binding to V-domain are solvent-exposed and accessible for the interaction. For S100A12 that binds to RAGE at the C1 domain, the available hexameric crystal structure shows a completely different arrangement of the dimmers. The interacting surface of S100P and S100B are inward-facing and not fully accessible for the interaction.

In conclusion, our data show that all the investigated S100 proteins increase cell proliferation in the investigated cancer cell lines. They trigger the activation of RAGE by binding to at least one of the three extracellular domain of Rage. In most cases, the interaction involves the V-domain, although S100A12 shows a different preference. The region between helix 3 and 4 of S100 proteins and the surface formed by strands 2, 5, 6 and some interconnecting loops seem to play a leading role in ligand recognition and in the subsequent receptor activation.

References

1. Sparvero, L.J., Asafu-Adjei, D., Kang, R., Tang, D., Amin, N., Im, J., Rutledge, R., Lin, B., Amoscato, A.A., Zeh, H.J. & Lotze, M.T. RAGE (Receptor for Advanced Glycation End Products), RAGE ligands, and their role in cancer and inflammation. *J. Transl. Med.* **7,** 1479-1500 (2009).
2. Schmidt, A. M., Vianna, M., Gerlach, M., Brett, J., Ryan, J., Kao, J., Esposito, C., Hegarty, H., Hurley, W. & Clauss, M. Isolation and characterisation of two binding proteins for advanced glycosylation end products from bovine lung which are present on the endothelial cell surface. *J. Biol. Chem.* **267,** 14987-14997 (1992).
3. Hori, O., Brett, J., Slattery, T., Cao, R., Zhang, J., Chen, J.X., Nagashima, M., Lundh, E.R., Vijay, S. & Nitecki, D. The receptor for advanced glycation end products (RAGE) is a cellular binding site for amphoterin. Mediation of neurite outgrowth and co-expression of RAGE and amphoterin in the developing nervous system. *J. Biol. Chem.* **270,** 25722-25761 (1995).
4. Yan, S.F., Ramasamy, R. & Schmidt, A.M. Mechanisms of disease: Advanced glycation end products and their receptor in inflammation and diabetes complications. *Nat. Clin. Pract. Endocrinol. Metab.* **4,** 285-293 (2008).
5. Schmidt, A.M., Hoffman, M., Taguchi, A., Yan, S.D. & Stern, D.M. RAGE, a multiligand receptor contributing to cellular response in diabetic vasculopathy and inflammation *Semin. Thromb. Hemost.* **26,** 485-493 (2000).
6. Schmidt, A., Yan, S. & Stern, D. The multiligand receptor RAGE as a progression factor amplifying immune and inflammatory responses. *J. Clinl. Invest.* **108,** 949-955 (2001).
7. Clynes, R., Moser, B., Yan, S., Ramasamy, R., Herold, K., Schmidt, A. Receptor for AGE RAGE): weaving tangled webs within the inflammatory respone. *Current Molecular Medicine.***7,** 743-751 (2007).
8. Bartling, B., Hofmann, H.S., Weigle, B., Silber, R.E., Simm, A. Down regulation of the receptor for advanced glycation end products (RAGE) supports non-small cell lung carcinoma. *Carcinogenesis* **26,** 293-301 (2005).
9. Englert, J.M., Hanford, L.E., Kaminski, L., Tobolewski, J.M., Tan, R.J., Fattman, C.L., Ramsgaard, L., Richards, T.J., Loutaev, I., Nawroth, P.P., Kasper, M., Bierhaus, A., Oury, T.D. A role for the receptor of advanced glycation end products in idiopathic pulmonary fibrosis. *Am. J. Pathol.* **172,** 583-591 (2008).

10. Yan, S., Ramasamy, R., Schmidt, A. Receptor for AGE (RAGE) and its ligands- cast into leading roles in diabetes and the inflammatory response. *J. Mol. Med.* **87,** 235-247 (2009).
11. Yan, S. D., Chen, X., Fu, J., Chen, M., Zhu, H., Roher, A., Slattery, T., Zhao, L., Nagashima, M., Morser, J., Migheli, A., Nawroth, P., Stern, D. & Schmidt, A. M. RAGE and amyloid-beta neurotoxicity in Alzheimer's disease. *Nature.* **382,** 685-691 (1996).
12. Christoff, G., Astrid, R., Durchdewald, M., Meth, J. N., Furstenberger, G. F., Decker, K., M., Enk, A., Arnold, B., Bierhaus, A., Nawroth, P.P., Hess, J. & Angel, P. RAGE signalling sustains inflammation and promotes tumour development. *J. Exp. Med.* **205,** 275-285 (2008).
13. Demling, N., Ehrhardt, C., Kasper, M., Laue, M., Knels, L., Rieber, E. Promotion of cell adherence and spreading: a novel function of RAGE, the highly selective differentiation marker of human alveolar epithelial type I cells. *Cell Tissue Res.***323,** 475-499 (2006).
14. Kuniyasu, H., Oue, N., Wakikawa, A., Shigeishi, H., Matsutani, N., Kuraoka, K., Ito, R., Yokozaki, H., Yasui, W. Expression of receptors for advanced glycation end products (RAGE) is closely associated with the invasive and a metastatic activity of gastric cancer. *J. Pathol.* **196,** 163-170 (2002).
15. Bhawal, U.K., Ozaki, Y., Nishimura, M., Sugiyama, M., Sasahira, T., Nomura, Y., Sato, F., Fujimoto, K., Sasaki, N., Ikeda, M.A., Tsuji, K., Kuniyasu, H. & Kato, Y. Association of expression of receptor for advanced glycation end products and invasive activity of oral squamous cell carcinoma. *Oncology.* **69,** 246-255 (2005).
16. Takada, M., Koizumi, T., Toyama, H., Suzuki, Y. & Kuroda, Y. Differential expression of RAGE in human pancreatic carcinoma cells. *Hepatogastroenterology* **48,** 1577-1580 (2001).
17. Kuniyasu, H., Yano, S., Sasaki, T., Sasahira, T., Sone, S. & Ohmori, H. Colon cancer cell derived high mobility group 1/amphoterin induces growth inhibition and apoptosis in macrophages. *Am. J. Pathol.* **166,** 751-760 (2005).
18. Neeper, M., Schmidt, A.M., Brett, J., Yan, S.D., Wang, F., Pan, Y.C., Elliston, K., Stern, D. & Shaw, A. Cloning and expression of a cell surface receptor for advanced glycosylation end products of proteins. . *J. Biol. Chem.* **267,** 14998-15004 (1992).

19. Dattilo, B.M., Fritz, G., Leclerc, E., Vander-Cooi, C.W., Heizmann, C.W. & Chazin, W.J. The extracellular region of the receptor for advanced glycation end products is comprised of two independent structural units. *Biochemistry.* **46,** 6957-6970 (2007).
20. Leclerc, E., Fritz, G., Vetter, S. W. & Heizmann, C.W. Binding of S100 proteins to RAGE: an update. *Biochim. Biophys. Acta.* **1793,** 993-1007 (2009).
21. Donato, R. Intracellular and extracellular roles of S100 proteins. *Microsc. Res. Technol.* **60,** 540-551 (2003).
22. Schaefer, B.W. & Heizmann, C.W. The multifunctional S100 protein family. *Trends. Biochem. Sci.* **21,** 134-140 (1996).
23. Schaefer, B.W., Wicki, R., Engelkamp, D., Mattei, M.G. & Heizmann, C.W. Isolation of a YAC clone covering a cluster of nine S100 genes on human chromosome 1q21:rationale for a new nomenclature of the S100 protein family. *Genomics.* **25,** 638-643 (1995).
24. Marenholz, I., Zina, M., Fischer, D.F., Backendorf, C., Zeigler, A. & Mischke, D. Identification of human epidermal differentiation complex (EDC)-encoded genes by subtractive hybridisation of entire YACs to a gridded keratinocyte cDNA library. *Genome Res.* **11,** 341-355 (2001).
25. Schutte, B.C., Carpten, J.D., Forus, A., Gregory, S.G., Horii, A. & White P.S. Report of the sixth international workshop on human chromosome 1 mapping 2000. *Cytogenet. Cell Genet.* **92,** 23-48 (2001).
26. Kisiel, L.S., Dempsey, A.C. & Shaw, G.S. Calcium dependent and independent interactions of the S100 protein family. *Biochem. J.* **296,** 201-214 (2006).
27. Bode, G., Luken, A., Kerkhoff, C., Roth, J., Ludwig, S. & Nacken, W. Interaction between S100A8/A9 and annexin A6 is involved in the calcium induced cell surface exposition of S100A8/A9. . *J. Biol. Chem.* **283,** 31776-31784 (2008).
28. Novitskaya, V., Grigorian, M., Kriajevska, M., Tarabykina, S., Bronstein, I., Berezin, V., Bock, E. & Lukanidin, E. Oligomeric forms of the metastasis related Mts1 (S100A4) protein stimulate neuronal differentiation in cultures of rat hippocampul neurons. *J. Biol. Chem.* **275,** 41278-41286 (2000).
29. Moroz, O.V., Burkitt, W., Wittkowski, H., He, W., Ianoul, A., Novitskaya, V., Xie, J., Polyakova, O., Lednev, I.K., Shektman, A., Derrick, P.J., Bjoerk, P., Foell, D. & Bronstein, I.B. Both Ca^{2+} and Zn^{2+} are essential for S100A12 oligomerisation and function. *BMC Biochemistry.* **10,** 1471-1489 (2009).

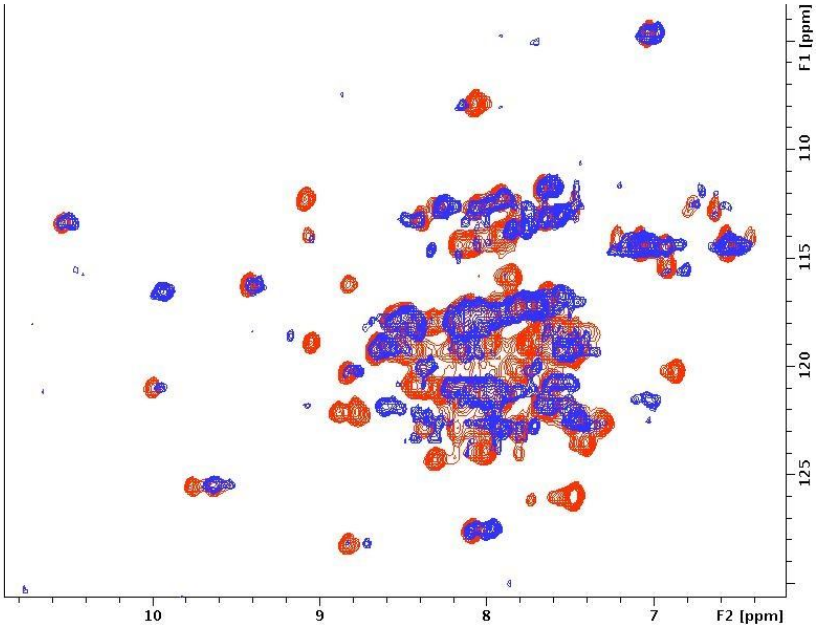
Heizmann, C.W. & Cox, J.A. New perspectives on S100 proteins: a multifunctional Ca2+-, Zn2+- and Cu2+-binding protein family. *Biometals.* **11,** 383-387 (1998).

30. Arnesano, F., Banci, L., Bertini, I., Fantoni, A., Tenori, L., Viezzoli, M.S. Structural interplay between calcium(II) and copper(II) binding to S100A13 protein. *Angew. Chem. Int. Engl.* **44,** 6341-6344 (2005).
31. Baudier, J. & Cole, R.D. Interactions between microtubule-associated tau proteins and S100B regulate tau phosphorylation by the Ca^{2+}/calmodulin-dependent protein kinase II. *J. Biol. Chem.* **263,** 5876-5883 (1988).
32. Millward, T.A., Heizmann, C.W. Schaefer, B.W. & Hemmings, B.A. Calcium regulation of Ndr protein kinase mediated by S100 calcium binding proteins. *EMBO J.* **17,** 5913-5922 (1998).
33. Donato, R. Calcium-independent, pH-regulated effects of S100 proteins on assembly-disassembly of brain microtubule protein in vitro. *J. Biol. Chem.* **263,** 106-110 (1988).
34. Baudier, J., Bergeret, E., Bertacchi, N., Weintraub, H., Gagnon, J. & Garin, J. Interaction of bHLH transcription factors with calcium-binding calmodulin and S100a proteins. *Biochemistry.* **34,** 7834-7846 (1995).
35. Zimmer, D.B. & Landar, A. Analysis of S100A1 expression during skeletal muscle and neuronal cell differentiation. *J. Neurochem.* **64,** 2727-2736 (1995).
36. Selinfreund, R.H., Barger, S.W., Pledger, W.J. & Van Eldik, L.J. Neurotrophic protein S100 beta stimulates the glial cell proliferation.
37. Donato, R. S100: a multigenic family of calcium modulated proteins of the EF-hand type with intracellular and extracellular functional roles. *Int. J. Biochem. Cell Biol.* **33,** 637-668 (2001).
38. Fuentes, M.K., Nigavekar, S.S., Arumugam, T., Logsdon, C.D. Schmidt, M., Park, J.C. & Huang, E.H. RAGE activation by S100P in colon cancer stimulates growth, migration and cell signalling pathways. *Diseases of the Colon and Rectum.* . **50,** 1230-1240 (2007).
39. Fuentes, M.K., Nigavekar, S.S., Arumugam, T., Logsdon, C.D. Schmidt, M., Park, J.C. & Huang, E.H. RAGE activation by S100P in colon cancer stimulates growth, migration and cell signalling pathways. *Diseases of the Colon and Rectum.* **50,** 1230-1240 (2007).
40. Arumugam, T., Simeone, D.M., Schmidt, A.M. & Logsdon, C.D. S100P stimulates cell proliferation and survival via receptor for advanced glycation end products (RAGE).). *J. Biol. Chem.* **279,** 5059-5065 (2003).

41. Chuong, C., Katz, J., Pauley, K.M., Bulosan, M. & Cha, S. RAGE expression and NF-κB activation attenuated by extracellular domain of RAGE in human salivary gland cell line. *J. Cell. Physiol.* **10,** 430-434 (2009).

3.4 Paramagnetic Probe attachment to S100 proteins for interaction studies

Two mutations, Q65G and K72E in EF-2 of S100A5 (based on sequence alignment with calbindin) could successfully convert the calcium binding site to a lanthanide binding site. The Ca^{2+} at EF-2 of this mutant protein could be selectively substituted with lanthanides.



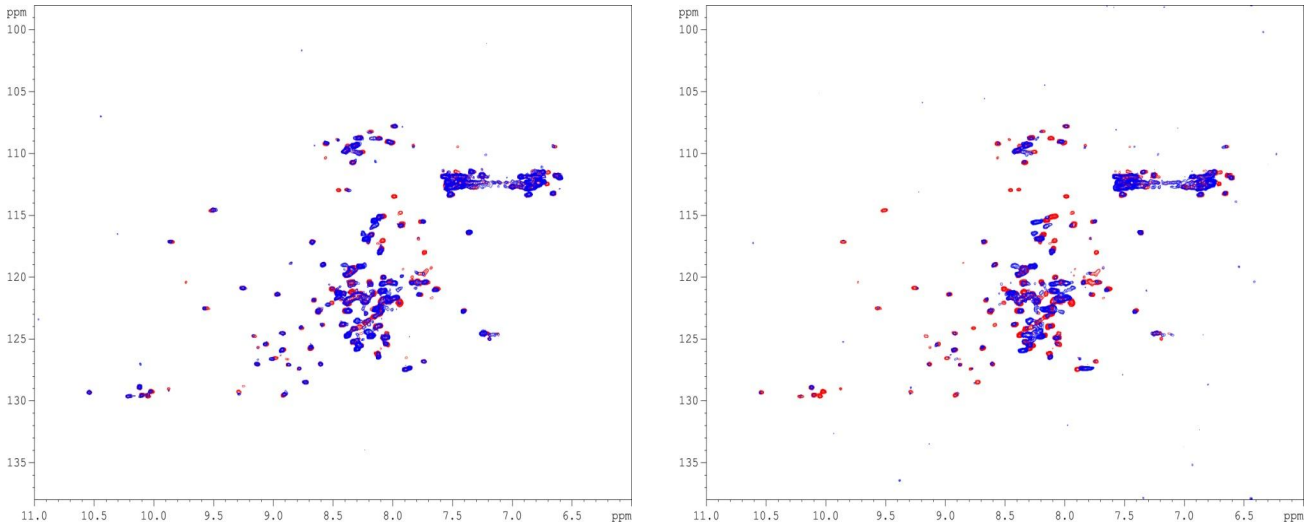
Red: 15N HSQC of Ca^{2+} S100A5 **Blue:** 15N HSQC of 1 Ca^{2+} 1Yb^{3+} S100A5

Temp: 298 K Spectra taken at 700 MHz

Figure 1: Overlay of Ca^{2+} S100A5 with 1 Ca^{2+} 1 Yb^{3+} S100A5 mutant

This lanthanide derivative of S100A5 was used to study the interaction with the V-domain of RAGE.

A B



15N V-RAGE 15N V-RAGE

15N V-RAGE + unl S100A5 mut (1Ca^{2+} 1Lu^{3+}) 15N V-RAGE + unl S100A5 mut (1Ca^{2+} 1Yb3+)

Temp: 298K Spectra taken at 700 MHz S100:RAGE 2:1

Figure 2: Overlay of HSQC of V-RAGE with the same titrated with A) Ca^{2+} S100A5 mutant and B) 1 Ca^{2+} 1 Yb^{3+} S100A5 mutant.

Titration of RAGE with Yb^{3+} S100A5 caused the disappearance of several peaks from the spectrum. Moreover, no significant shifts or appearance of new peaks of the complex due to the paramagnetic effect could be detected. As such this lanthanide derivative could not be used further for structural characterisation of the complex.

3.5

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] *Luca Sgheri,*[4] *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.

[4]Istituto per le Applicazioni del Calcolo, Sezione di Firenze, CNR, Via Madonna del Piano 10, 50019 Sesto Fiorentino, Italy

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 X-ray scattering (SAXS) 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 ^{1}H-^{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 ^{1}H 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 ^{1}H-^{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 disulphide 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 starts 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. *We are now checking whether this can also be shown by comparing the results of the 3+0 calculations with the 2+1 calculations.*

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.

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.

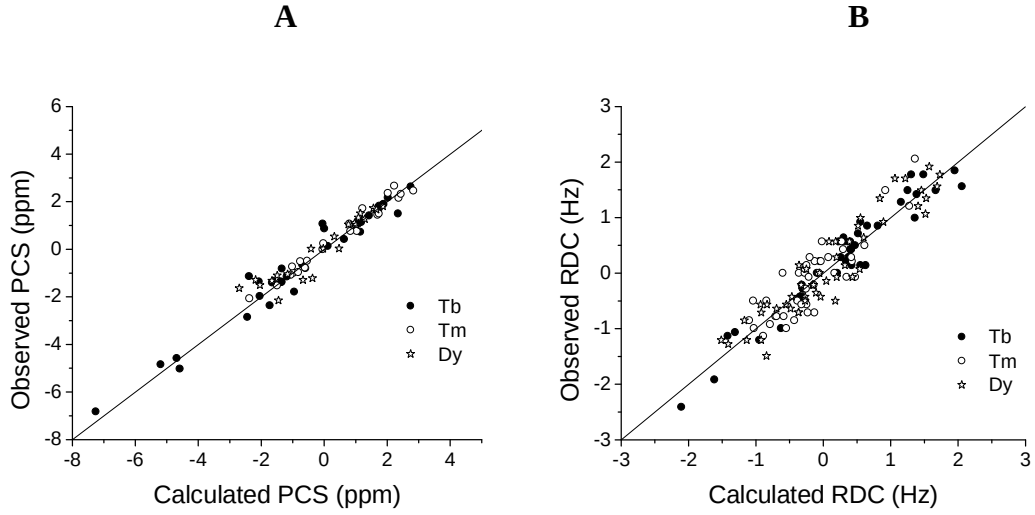


Figure 2 Superimposition of a region of the ^{1}H-^{15}N HSQC spectra of Lu^{3+} (red), Yb^{3+} (green) and Tm^{3+} (blue) ClaNP-5 CaM.

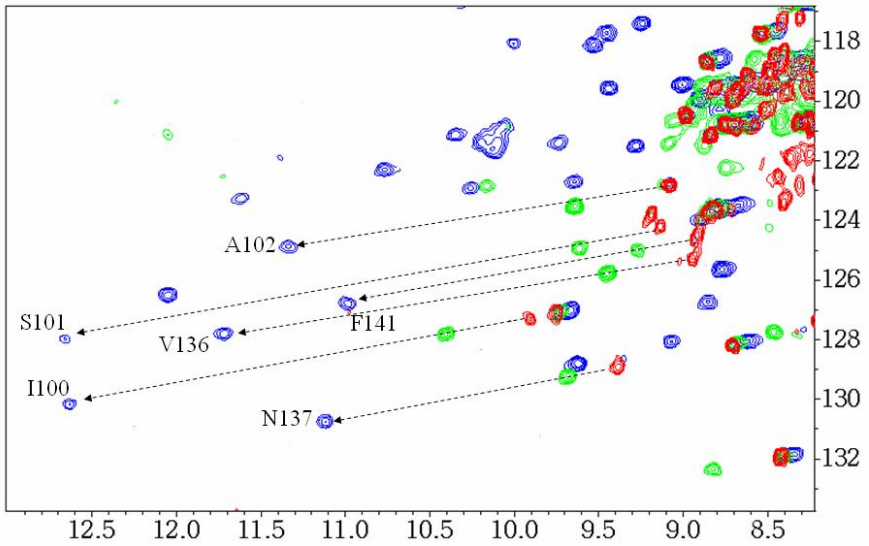


Figure 3 (A) Observed versus calculated pcs of C-terminal domain nuclei for the Yb^{3+} (●) and Tm^{3+} (○) ClaNP-5 Ca_4CaM. **(B)** Observed versus calculated rdc of C-terminal domain HN for the Yb^{3+}-ClaNP-5 Ca_4CaM. The solid symbols (●) indicate the values calculated from the best fit parameters ($\Delta\chi_{ax}$= 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.

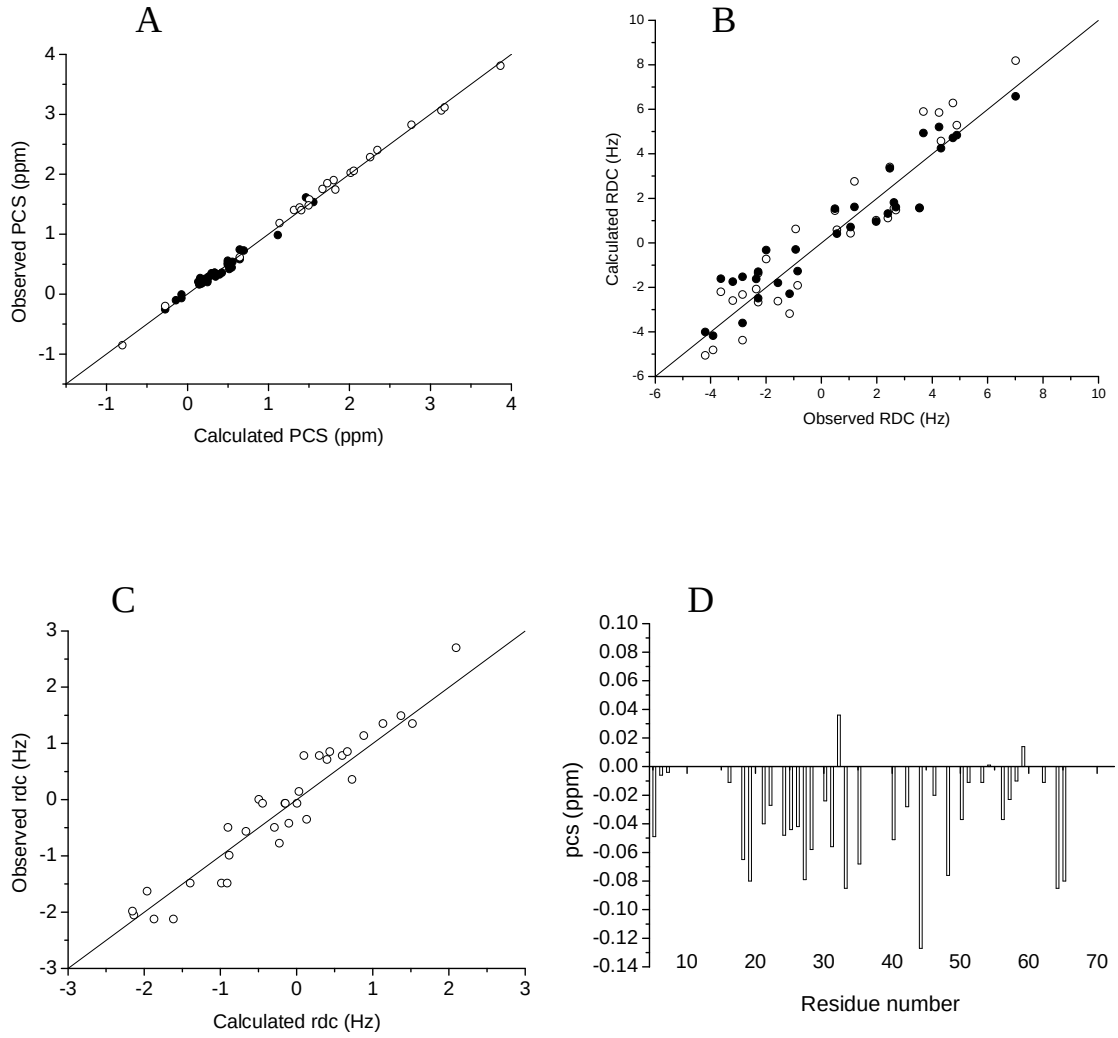


Figure 4. The lanthanide ions (in blue) are placed at a distance of 8.3 Å and 6.5Å from the C^{α} atoms of residues Cys-107 and Cys-111. The orientations of the magnetic susceptibility anisotropy tensors are shown for the Yb^{3+} (magenta) and Tm^{3+} (brown) metals. The *z* axes of the anisotropy tensors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

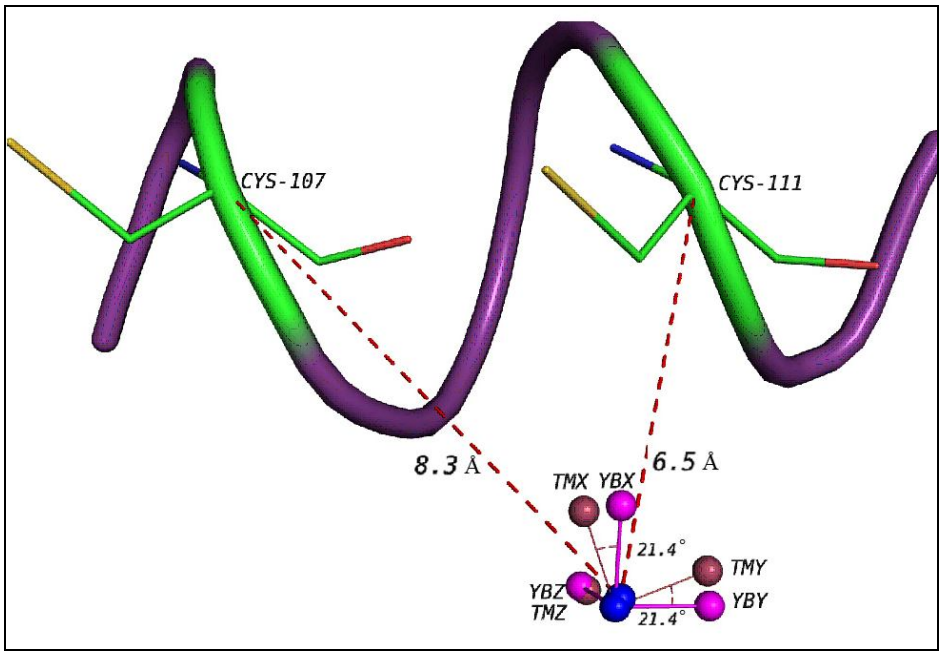


Figure 5. Arbitrary orientation tensors centered in the center of mass of the C-terminal domain, color-coded with respect to the MO of the corresponding conformation from blue (lower than 5%) to red (greater than 30%) for 400 structures generated randomly with RANCH. Panel **A** shows the results obtained 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.

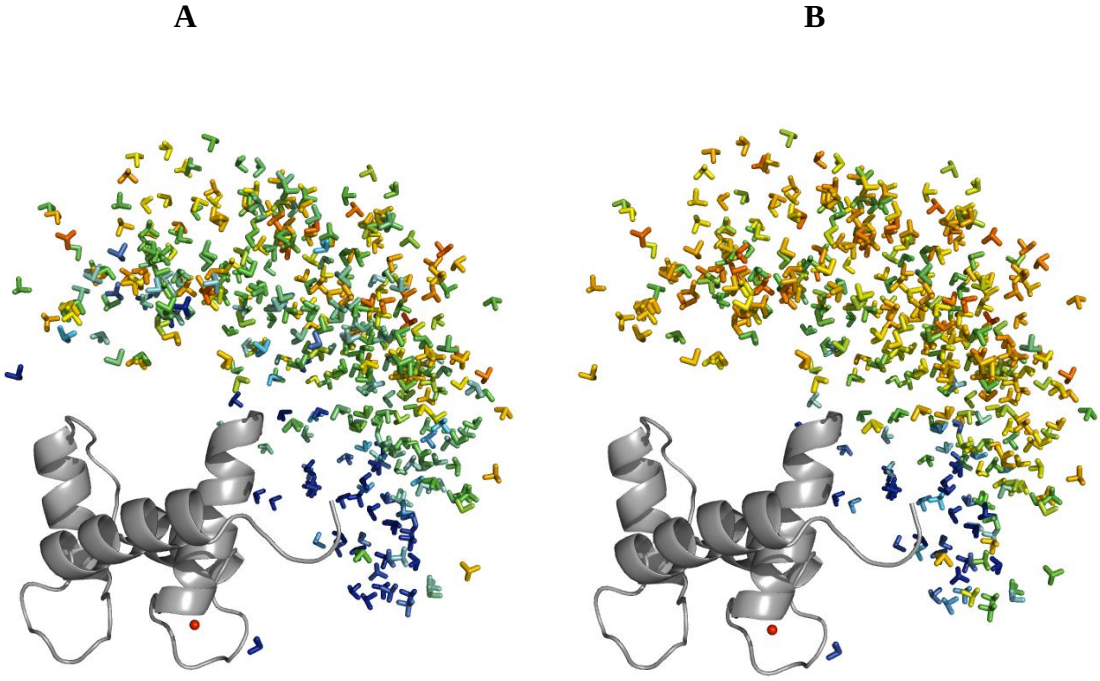


Figure 6. Number of conformations as a function of the MO values calculated from the pcs and rdc arising with metals in the N-terminal domain (3+0 case) and by including pcs and rdc of Tm^{3+}-ClaNP-5 Ca_4CaM (3+1 case).

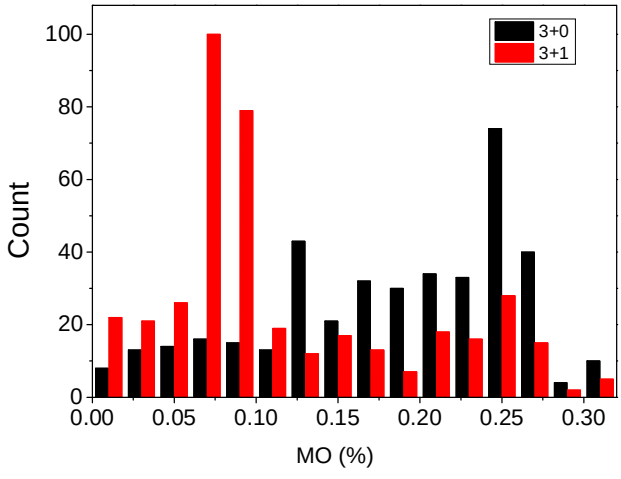


Figure 7. TF (describing the best agreement between any conformational ensemble and the experimental data) as a function of the weight of the conformation for which the MO is calculated. Each curve corresponds to a different conformation. The MO is defined by the intersection between the TF curve and 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.

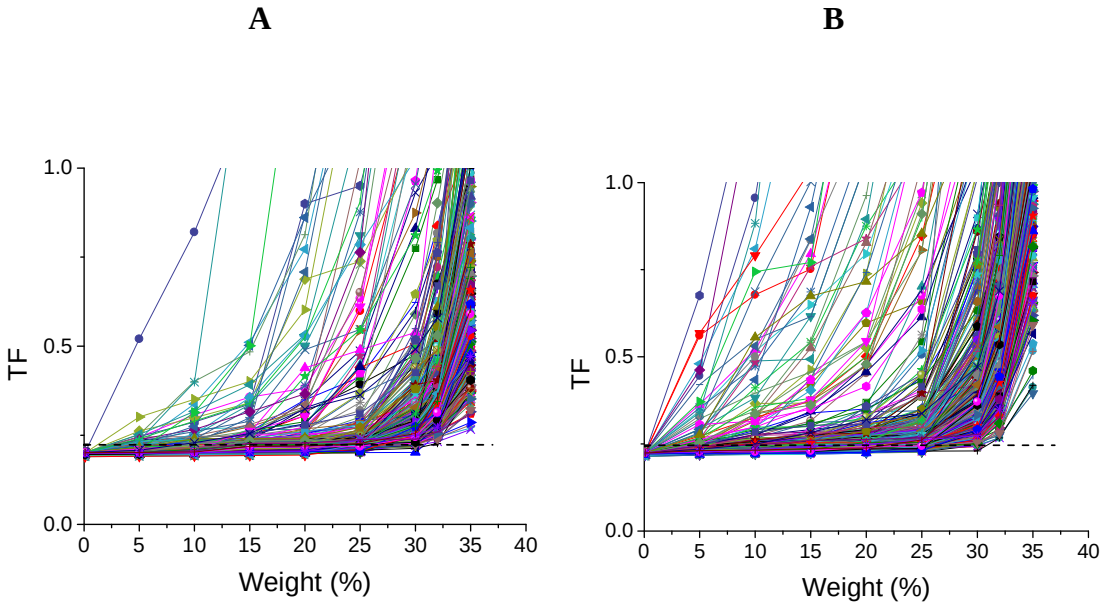
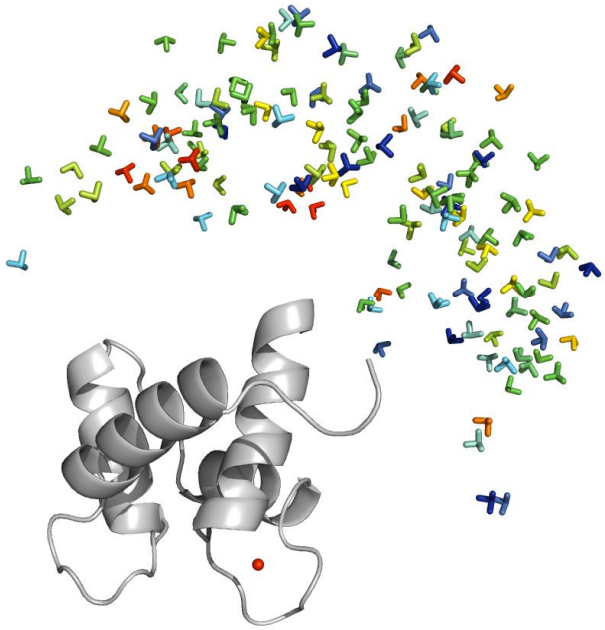


Figure 8. Arbitrary orientation tensors centered in the center of mass of the C-terminal domain, color-coded with respect to the difference in MO values upon inclusion of pcs and rdc of Tm^{3+}-ClaNP-5 CaM. The colors change from blue (difference in MO of 0.1) to red (difference in MO of 0.25) for the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



References

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*, 41-247.
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.

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

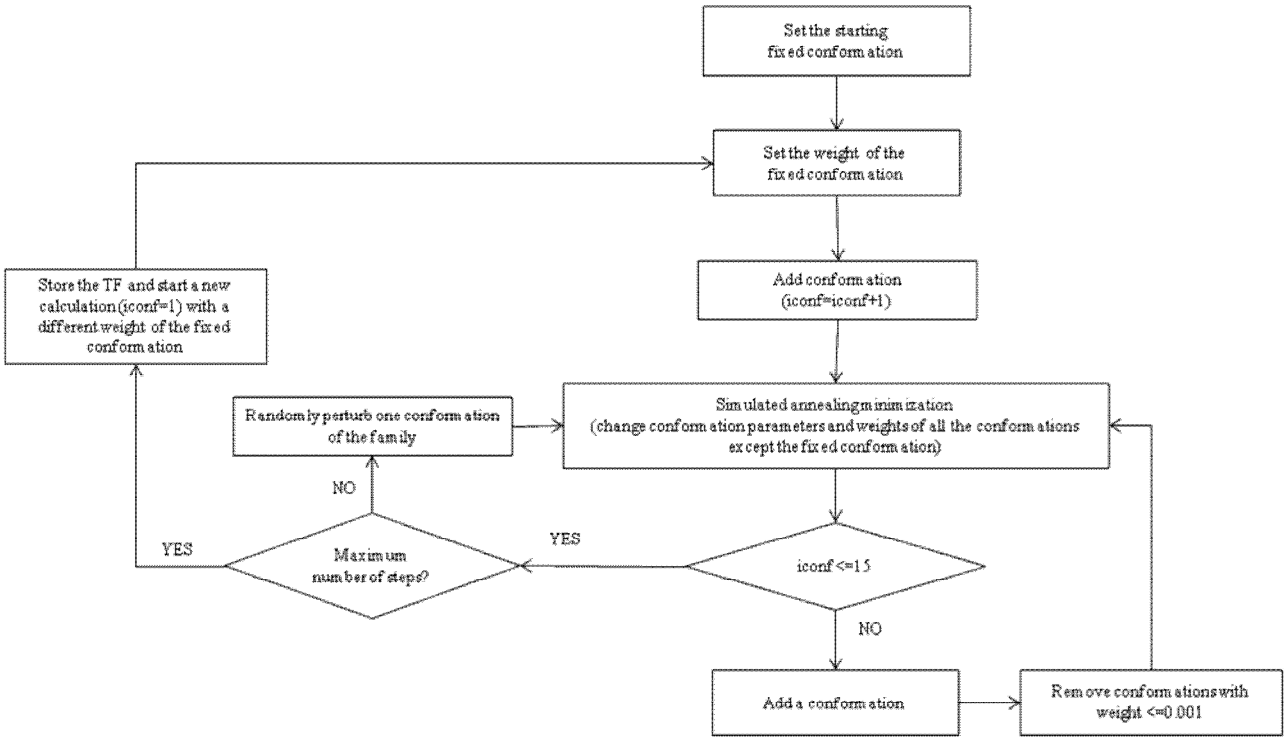
conformations with different weights (the sum of all weights being constrained to 1) were then added one by one, and each time a simulated annealing minimization was performed, in order to obtain the best possible agreement with the experimental data. A maximum number of 15 conformations was used to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

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

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:

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

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

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

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

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.

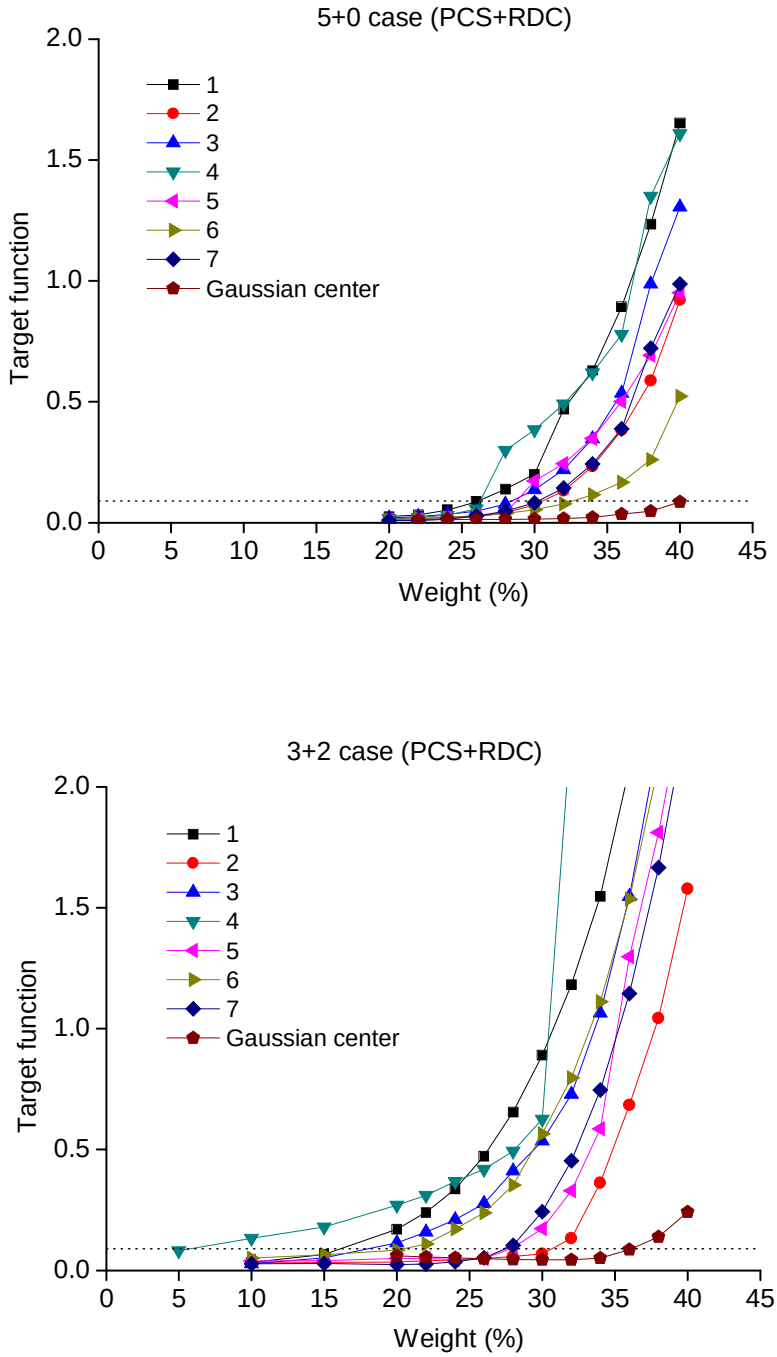
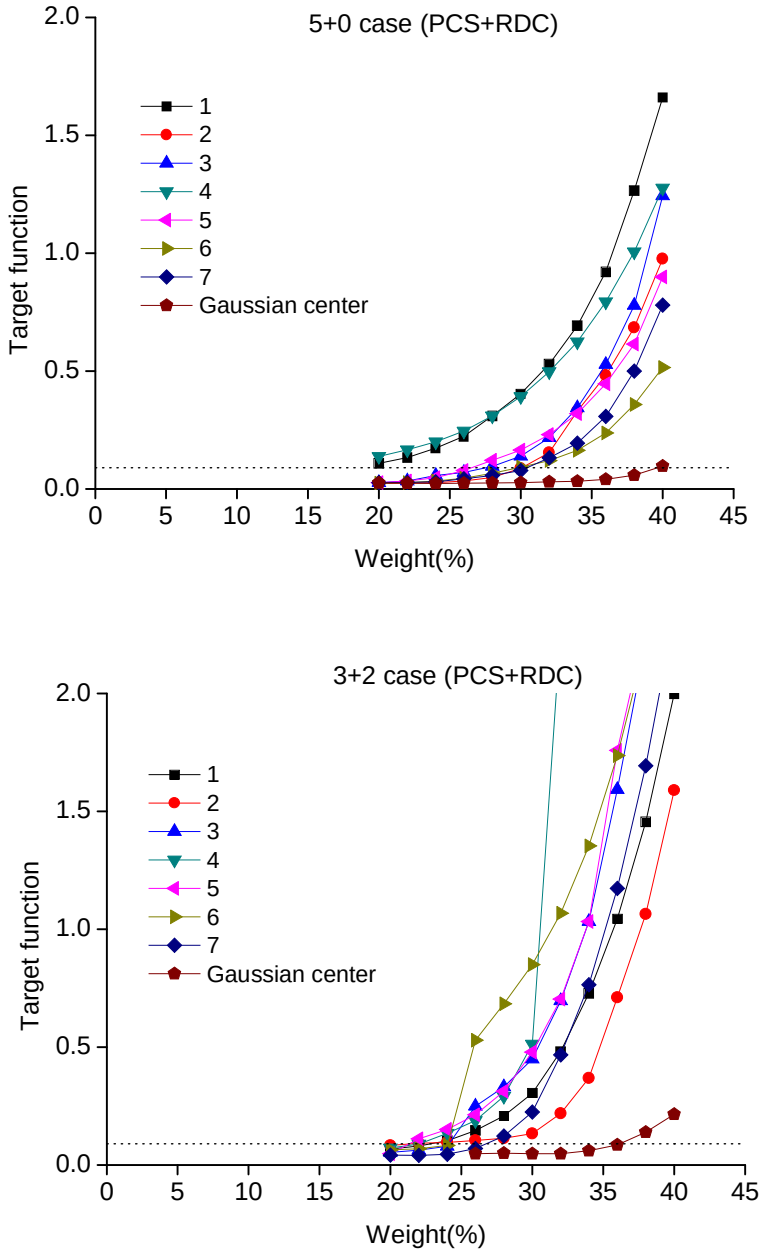
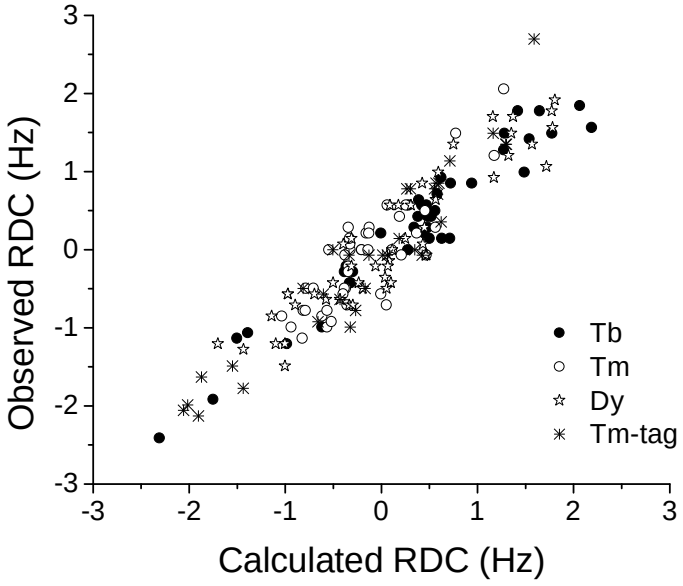


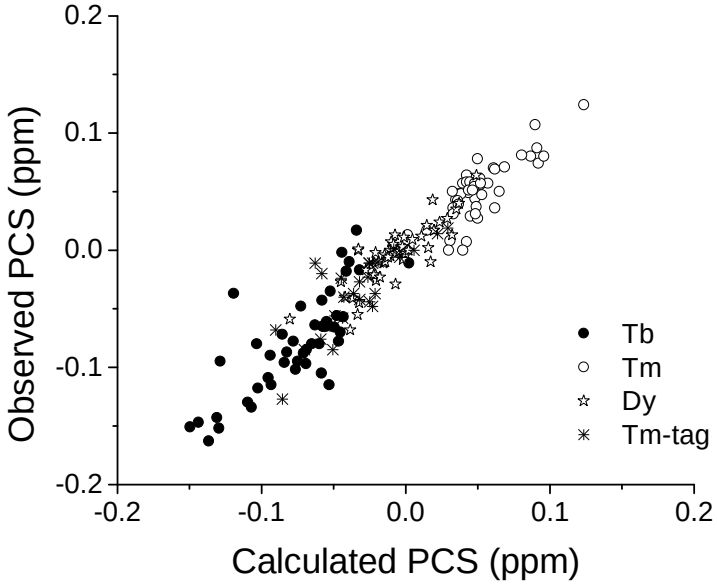
Fig. S2.3. MO analysis of the eight selected conformations within the simulated Gaussian distribution . Pcs and rdc data are simulated with addition of a Gaussian error.



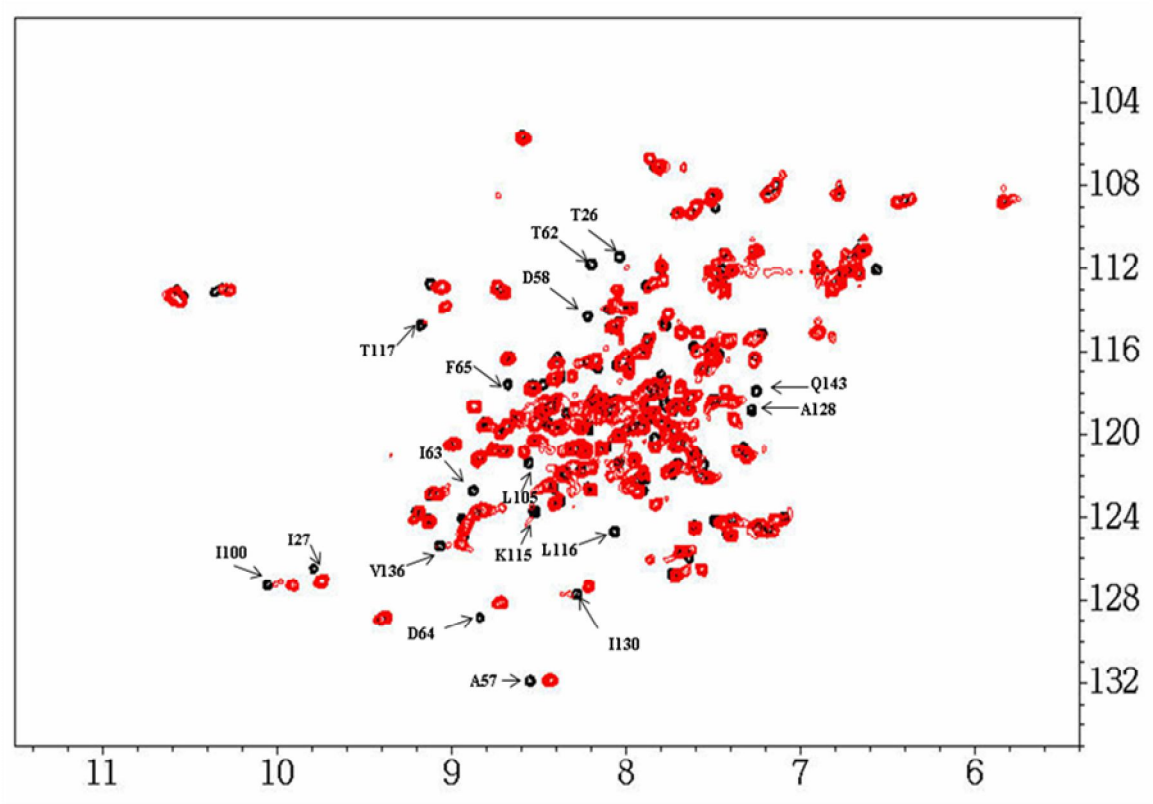
S3: Best fit graphs

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
93	ASP	-2.412	2.058	-1.277

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
95	ASP	3.477	21.853

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

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

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
85	ILE	-	0.124	-
86	ARG+	-	-	0.013
89	PHE	-0.095	0.070	0.002

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
65	PHE	-0.009	-
68	PHE	-0.014	-
70	THR	-0.029	-

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
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

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
145	MET	0.402	2.552

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

4. Conclusions

The solution structure of S100A5 was determined for apo and Ca^{2+} forms by NMR spectroscopy. Both forms are homodimeric in solution. The overall structures of apo and Ca^{2+} S100A5 are in good agreement with those obtained for other S100 proteins such as S100A1, S100A4, S100A6, S100A8, S100A12, S100B. Comparison of the apo and calcium-loaded structures shows that the N-terminal EF-hands are similar to one another, indicating there is no large conformational rearrangement in this region upon calcium binding. In contrast, the C-terminal EF-hand undergoes a major conformational change upon calcium binding. This rearrangement includes a considerably different orientation of helix III and non-negligible changes in helix IV and the hinge loop. These rearrangements upon calcium binding are similar to those observed for other S100 proteins with the exception of S100A10 which has a 'calcium ready state' in N-and C-terminal EF-hands although it does not bind calcium. In S100A5, the angles between helices III and IV changes from 168° to 118° on passing from the apo to calcium form. The structural changes induced by Ca^{2+} binding to the dimer leads to the exposure of two symmetrically positioned clefts defined by helix III, helix IV, the hinge loop and the last C-terminal residues, in a way similar to that found in other S100 proteins, where target proteins can be accommodated. These changes are important as they must be related to the functions of this protein which are still unknown. For the first time, relaxation measurements were performed on apo and calcium forms of a S100 protein. A large mobility was observed in the hinge loop which is not quenched in the calcium form. The structural changes induced upon calcium binding were found to change the global shape and distribution of hydrophobic and charged residues of the S100A5 dimer in a modest but significantly different manner with respect to the closest homologues S100A4 and S100A6.

Despite sharing a very high degree of structural homology, S100 proteins exhibit diverse functional roles. Several S100 proteins are secreted into the extracellular space. They exert their cellular effect through activation of the cell surface receptor, RAGE. However, the mechanism of activation varies among the different S100 proteins. Therefore, S100-RAGE interaction was investigated using a set of 7 recombinant S100 proteins (some reported to be extracellular and others not reported). Studies were based on cell line and NMR experiments. It was found that all these S100 proteins activate RAGE by increasing cell proliferation but with different potencies. For each protein, the concentration with maximal effect on cell proliferation was determined. S100A4 and S100A5 were found to have optimal influence on cell proliferation at a concentration 10 times lower compared to the other S100 proteins. NMR experiments revealed that no all S100 proteins bind to the

same extracellular domain of RAGE. S100A12 binds to the C1C2 domain while others bind to the V-domain. These findings were further verified through competition experiments in cell lines where cells were treated with S100 proteins in the presence of different constructs of extracellular RAGE, V, VC1 and C1C2. The RAGE domain which binds to each S100 protein, thereby preventing activation of the receptor on the cells and hence cell proliferation was determined. Our results from cell lines were consistent with NMR experiments for all the S100 proteins studied. The interaction of S100P with RAGE was investigated in detail and a model was proposed of the complex base on chemical shift perturbations. A dimer of S100P was found to interact with a monomer of VC1-RAGE. For the first time, a study was conducted on 7 different S100 proteins under the same conditions to avoid discrepancies arising from previous studies. Our studies also shed light on the possible biological function of S100A5. Although this protein is not reported to be extracellular, it was found to interact with the V-domain of RAGE in a manner comparable to other S100 proteins which are known to be secreted into the extracellular space. Thereby, our finding provides a hint for the extracellular role of S100A5 whose functional significance is yet to be explored. Moreover, the fact that S100-RAGE interactions promote cell proliferation suggests their role in tumour development. Hence these complexes can be significant drug targets.

Till now the probable spatial conformations of the two-domain protein CaM were studied with paramagnetic ions in the N-domain, i.e. N-domain as fixed and the C-domain as moving. Studies were performed using a paramagnetic probe at the C-domain, whereby the N-domain is moving. Combined analysis of these two sets of data show that the simultaneous use of paramagnetic restraints arising from lanthanide ions located in different domains of proteins experiencing interdomain mobility is quite informative for the determination of the maximum occurrence of any conformation. The information obtained from paramagnetic ions located in different domains of a protein is more than that obtained by placing the same number of metal ions in the same site mostly due to differences in orientations of the magnetic susceptibility anisotropy tensors. In case of CaM, we have shown that the maximum occurrence of several conformations is considerably reduced by the addition of pcs and rdc arising from a single metal ion rigidly attached to the C-terminal domain in comparison with the values obtained with three metal ions placed in the N-terminal domain of the protein. As a result, the probable conformations experienced by the protein can be more accurately mapped. The significance of this study also lies in the broad applicability of the methodology. It can serve as a general approach to study any multi-

domain protein or two interacting protein partners where there is mobility between the domains or the partners.

Publications

This thesis was based on the following publications:

1. Solution structure and dynamics of S100A5 in the apo and Ca^{2+} bound states. Bertini, I., **Das Gupta, S.**, Hu, X., Karavelas, T., Luchinat, C., Parigi, G., Yuan, J. ***J. Biol Inorg Chem*** (2009) 14:1097-1107.
2. New insights into the mechanism of RAGE activation by S100 proteins. Bertini, I., Borsi, V., **Das Gupta, S.**, Fragai, M., Luchinat, C., Melikan, M., Nesi, A. (*in preparation).*
3. Improving the maximum occurrence analysis of calmodulin conformations in solution by placing paramagnetic ions in both protein domains. Bertini, I., **Das Gupta, S.**, Hu, X., Keizers, P.H.J., Luchinat, C., Nagulapalli, M., Parigi, G., Sgheri, L., Ubbink, M. *(in preparation).*