University of Florence

International Doctorate in Structural Biology

Cycle XXII (2007-2009)

Intermolecular/interdomain interactions in proteins

Ph.D. thesis of

Valentina Borsi

Tutor

Prof. Giacomo Parigi

Coordinator

Prof. Claudio Luchinat

S.S.D. CHIM/03

This thesis has been approved by the University of Florence, the University of Frankfurt and the Utrecht University

Table of contents

Chapter 1

Introduction

1.1 Covalent and weak interactions in proteins

Life processes consist of building up and breaking down molecules and the knowledge of the nature of chemical bonds that link atoms together in a molecule is important to understand biological processes[1]. The combination of weak interactions, together with covalent bonds, determines the structure of proteins and drive the protein-protein interaction processes.

The majority of bonds in an organic molecule are covalent; these are formed when partially occupied orbitals of interacting atoms overlap, that is pairs of electrons are shared between atoms. Such bonds are strong and of short range. When a bond does not involve the sharing of pairs of electrons, we referred to non-covalent bonds. In general non-covalent bonding address to weak interactions that occur at distances of several angstroms and originate due to the electrical properties of the subsystem. Usually with the term “non-covalent interactions” we refer to these as a whole, indeed often they work in concert, thus providing force to hold the molecule, or part of the molecule together. The most common are the van der Waals interactions, that are short range interactions, so that they occur at distances comparable with the size of the interacting atoms, thus involving only neighbouring atoms. Such forces are about 100 times weaker than a covalent bond. Nevertheless, they tend to pull molecules together. When two molecules have complementary shapes hundreds or even thousands of v.d.W. forces may form between them. The Hydrogen bonds are interactions which are at the boundary between the covalent and non-covalent interactions. They take place between pairs of atoms only if one of them is a proton donor and the other one is a proton acceptor. This is due to the fact that an hydrogen atom has a significant partial positive charge if is covalently bound to a more electronegative atom, such as oxygen, and is attracted to a neighboring atom that has a significant partial negative charge. In contrast to the other two types, electrostatic interactions are long range ones. This means that electrostatic interactions are also relevant beyond the limits of the closest neighbours. Electrostatic forces are generated between the opposing charges and the molecules align themselves to increase the attraction.

Proteins are polymers of 20 different aminoacids bonded each other by covalent bonds (peptide bonds). The sequence of the different aminoacids constitutes the primary structure of the protein and determines how the protein folds into higher-level structures. The secondary structure can take the form either of alpha helices or of beta strands, formed through

hydrogen-bond interactions between NH and CO groups in aminoacids of the polypeptide backbone. Such elements of either alpha helix or beta sheet, or both, as well as loops and links that have no secondary structure, are folded into a tertiary structure. Finally, many proteins are formed by association of the folded chains of more than one polypeptide, thus constituting the quaternary structure of the protein[2].

Although the primary structure is due to the presence of covalent bonds, higher-level structures are principally due to non covalent, weak interactions. These interactions arise from several contributions: 1) van der Waals forces between the atoms forming the aminoacids; 2) hydrogen bonds between hydrophilic residues that are able to make such bonds to one another, to the peptide backbone, to polar organic molecules, and to water; 3) electrostatic interactions between ionizable groups of the aminoacids. Moreover it should be taken into account the contribution of the hydrophobic effect that arise from the tendency of hydrophobic aminoacids to avoid contact with water and to pack against each other.

1.2 Protein interaction network

Noncovalent forces control not only the folding of proteins but also their interactions with other proteins, nucleic acids, substrates, coenzymes and molecules in general. The comprehension of how the structure of the biomolecules affects their functions is one of the aim of structural biology. However, to really understand the cellular organization and the mechanism of the biological systems, not only the structure of the molecules and biomolecules are important but also the interaction among them. Moreover, protein flexibility and the dynamics of intermolecular interfaces can regulate binding affinity and specificity in molecular recognition, and can also have a profound effect on determining the thermodynamics and kinetics of the binding process.

1.2.1 Protein-protein interactions

Protein–protein interactions (PPIs) are known to be the major components of a wide variety of cellular events, they are operative at almost every level of cell function: in the structure of sub-cellular organelles, the transport across biological membranes, muscle contraction, signal transduction, regulation of gene expression, to make some examples. Complete sequencing of genomes has revealed that an organism can contain tens of thousands

of PPIs, thus forming all together a complex network[3,4].

Protein-protein interactions can be classified into different types depending on the similarity between the interacting subunits (homo- or hetero-oligomers), the lifetime of the complex, (weak or transient interactions) and the location of interacting partners within one or on two polypeptide chains. Finally, since proteins usually work in a crowded environment with many potential binding partners, most proteins can be very specific in their choice of partner, although others can be multispecific, having multiple (competing) binding partners on coinciding or overlapping interfaces[5]. Thus, regulation of cell function by the interactions of these proteins is balanced not only by the relative affinities of the various protein partners but also by the modulation of these affinities by the binding of ligands, other proteins, nucleic acids, ions such as Ca2+, and covalent modification, such as specific phosphorlyation or acetylation reactions. This results in different biological outcomes and is extremely important for many diseases. Indeed, aberrant protein-protein interactions are involved in a number of neurological disorders such as Creutzfeld-Jacob and Alzheimer's disease.

Therefore, PPI mapping would be of great value for understanding the molecular mechanism of cellular functions as well as diseases[6]. To do this, a structural characterization of the domains and of their dispositions within the protein and in their complexes with partners is fundamental. But in addition to this, we need a complete knowledge of the basis for specificity in these systems and comparative studies of similar interacting partners or mutated domains in order to bring to light the energetic properties linked to a particular sequence/structure.

1.2.2 Internal protein dynamics

To date, a number of three-dimensional protein structures have been solved at atomic resolution, nevertheless, explaining the protein function solely from the static three-dimensional structure is expected to be difficult.

Actually, proteins are dynamic over a spectrum of time scales and there is a deep correlation between dynamics and molecular function. Changes in conformational dynamics between folded and unfolded states contribute a significant entropy component to the energetics of protein stability, but protein dynamics play an important role also in molecular recognition processes (since entropic effects due to changes in internal dynamics can have a deep impact on binding affinities), in ligand binding (which often involves movement of molecules into areas that would normally be occluded), in catalysis (where conformational

rearrangements can juxtapose catalytic residues), and in allostery (where structural fluctuations can transmit information between different sites in a protein).

1.2.3 Protein-ligands interaction

Mapping interactions at protein-ligand binding sites is an important aspect of understanding many biological reactions and is a key part of drug design. It has been shown for a large numbers of targets that the 3D-structure of the protein can be used to design small molecules binding tightly to the protein[7,8] and future progress in this field can be achieved only with a complete understanding of the protein-ligand interactions. The majority of the currently available drugs act via noncovalent interactions with the target protein. Therefore, these forces are of particular interest in drug design.

The data now available on 3D-structures of protein-ligand complexes clearly indicate that there are several features found basically in all complexes: a high level of steric complementarity between the protein and the ligand (the lock-and-key paradigm); a high complementarity of the surface properties (lipophilic parts of ligand usually are found in contact with lipophilic parts of the protein, and the same usually occur with polar groups that form hydrogen bonds or ionic interactions); finally the ligand usually binds protein in an energetically favorable conformation.

The enthalpic and entropic components of the binding affinity can be determined experimentally and the available data indicate that there is always a substantial compensation between enthalpic and entropic contributions[9].

For many protein complexes only relatively small regions of the binding surface, often called “hot spots”, contribute a major part of the binding energy. The small ligands bind almost exclusively to well-defined, localized regions of proteins, independently of their affinity[10]. Once these hot spots are identified, binding interactions with adjacent regions of the protein surface can be subsequently explored to increase selectivity and improve affinity[11].

1.3 Metalloproteins and immunoglobulin cell surface proteins as models

Calcium-binding proteins and zinc proteases are two large classes of proteins involved in several physiological and pathological processes. In these proteins weak interactions play a

crucial rule in structural organization and target recognition.

EF-hand calcium-binding proteins are involved in the regulation of various cellular processes where they act as calcium sensors or as signal modulators (calcium buffers). The biological activity of these proteins rely on ion-induced conformational change or on features like ion binding affinity, selectivity, and kinetics. Matrix metalloproteinases are zinc-containing proteases involved in extracellular matrix degradation, which is a fundamental step in many physiological processes like tissue remodeling and repair.

1.3.1 EF-hand calcium binding proteins

The EF-hand family of Ca^{2+}-binding proteins (CaBPs) provides a rich framework for investigating fundamental relationships between weak interactions and biochemical function. The EF-hand motif is one of the simplest motifs with specific function and is among the most common in animal cells[12]. It consists of two α helices linked by a loop region that binds calcium and was first found in the structure of parvalbumin. It's specific for calcium-binding and, beside parvalbumin is present also in calmodulin, troponin-C and other calcium-binding proteins that regulate cellular activities towards the binding of the ion. These motifs are organized into structural units/domains containing two or more EF hands that form highly stable helical bundles [13].

The S100 human protein family consists of at least 25 calcium-binding proteins and constitute the largest family within the EF-hand protein superfamily. The nomenclature of S100 genes was firstly introduced by Schaefer et al. [14] in order to overcame the considerable confusion generated by the plethora of names given to the early members of the family. In contrast to the abundance of S100 genes in vertebrates, they are completely absent in invertebrates. Interestingly, 21 human genes of the family of S100 are clustered in region 1q21 of human chromosome 1. This chromosomal region exhibit several rearrangements which occur during tumor development, thus suggesting a correlation between S100 proteins with neoplasias[15-17].

S100 members are best known as mediator of intracellular Ca^{2+} signals, nevertheless certain members of this family are also secreted outside the cell, exerting extracellular actions in an endocrine, paracrine and autocrine manner[18].

Although the sequences of the family show some diversity, the key structural features of all S100 proteins are highly conserved[19] (Figure 1). S100 proteins contain two EF-hand motifs, one in the N-terminal domain (composed by helix I, loop I and helix II) and one in the

the C-terminal domain (composed by helix III, loop II and helix IV). The two domains are connected by a linker, called “hinge-loop”. The first N-terminal EF-hand is unconventional, because its loop is usually composed by 14 aminoacids; the second one, in the C-terminal domain, is canonical[20]. A consequence of the longer loop in the N- terminal EF-hand of S100 proteins is the different affinity of the binding of calcium(II) to individual EF-hand, due to the different coordination of calcium. The canonical C-terminal domain binds calcium in a similar manner to calmodulin and troponin-C resulting in a higher affinity site[18,21]. The N-terminal domain mainly binds the ion through main-chain carbonyl groups, in addition to the bidentate side chain of glutamate at the end of the loop, and this reduces the binding affinity up to 100 times[22].

With the exception of calbindin D9K, which is monomeric, all the other structures of the S100 proteins revealed a homo- and, in some case, hetero-dimerization, in which every monomer comprise two EF-hand motifs[15]. Certain members of the family also form active tetramers or larger oligomers. Monomers are related by a two-fold axis of rotation and the major contributors to the dimer interface are the helix I and IV of each monomer that are ordered in a X-type four-helix bundle. This relationship is maintained both in the apo- and in the calcium-bound states[18].

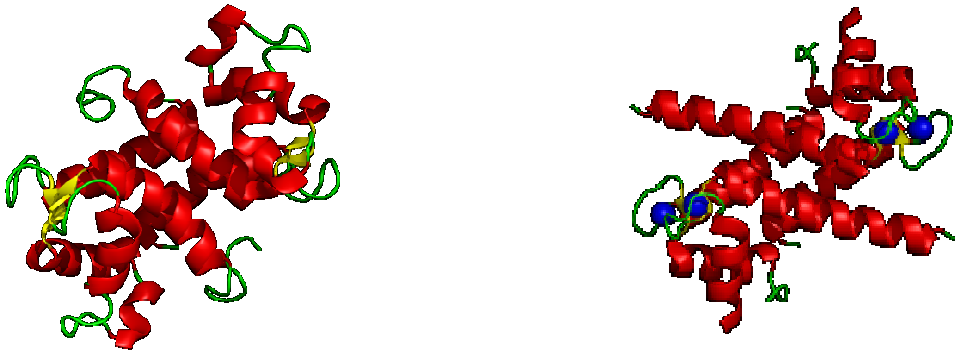


Fig. 1: Ribbon representation of the S100P dimer. (A) Solution structure of apo-S100P (PDB: 1OZO). (B) X-Ray structure of holo-S100P (PDB: 1J55). Calcium ions are depicted in blue.

Calcium binding to the S100s occurs in response to increases in intracellular calcium concentration. Upon calcium(II)-binding most of S100 proteins experiences a conformational change that involve helix III, which is antiparallel to helix IV in the apo state. As a consequence of the conformational rearrangement, the two helices become almost perpendicular. This movement “opens” the structure and exposes a wide hydrophobic cleft

that lays out as a binding site for targets[23]. Calcium binding to the N-terminal EF-hand, instead, causes only minor alterations of its backbone conformation. On the other hand, some family members, S100A10 and calbindin D9k, do not undergo changes in their conformation and have the function of calcium buffer and transport[24,25]. Moreover, not all the S100 interactions are dependent on binding to calcium, indeed, calcium-independent interactions have also been described. In most cases the partners of apo-S100s are enzyme (i.e. glycogen phosphorylase for S100B and S100A1[26] and transglutaminase for S100A10 and S100A11[27]), but the most important calcium-independent interactions of the S100 proteins are their abilities to form homo- and heterodimers, as well as some higher-order complexes.

Besides calcium, some S100 proteins have also been shown to bind zinc (i.e. S100B[28] S100A2[29], S100A7[30], S100A12[31]), however, because of its subnanomolar intracellular concentration, binding of Zn^{2+} in the cytoplasm is rather unlikely. However, zinc binding to S100 proteins may occur in the extracellular space where the Zn^{2+} concentration may be much higher locally for a short time (in the brain it can rise as high as several tens to hundreds of μM[32]). At this regard the zinc has been reported to modulate the interaction of S100B with tau protein[33].

According to their high specialization, S100 proteins are involved in the regulation of a variety of intracellular processes, such as protein phosphorylation (τ protein and p53 by S100B[33,34]), enzyme activities (phospholipase A_2 by S100A10[35]), Ca^{2+} homeostasis (S100A1 by directly affecting the ryanodine receptor[36]), cell growth and differentiation, transcription factors, cytoskeleton dynamics and the inflammatory response.

Secretion has been demonstrated for several members of the S100 protein family (i.e. S100B, S100A1, S100A2, S100A4, S100A6, S100A7, S100A8/A9, S100A12). The mechanism of secretion is still obscure, although, for the S100A8/A9 complex it is suggested a secretion pathways that depends on an intact microtubule network[37]. At the same time, for S100B it's known that secretion is not affected by the endoplasmic reticulum-Golgi classical secretion pathway inhibitor, brefeldin A. Despite the lack of information regarding the secretion mechanism of S100s, extracellular roles have been described for several S100 proteins. For examples S100A8/A9 heterodimer is chemotactic for mediating inflammation[38], S100B exhibit neurotrophic activity[39], and S100A4 has angiogenic effects[40].

Calmodulin (CaM) is one of the many Ca^{2+}-binding regulatory proteins. It's an ubiquitous, multifunctional protein that is widely distributed in nature and can bind at least 30 different targets with the subsequent regulation of several cellular processes, such as cellular

division and differentiation, gene transcription, ion transport by channels and muscle conctraction. It was discovered as the activator of the cyclic nucleotide phosphodiesterase in brain and heart [41,42], and is composed by two domains (N- and C-terminal) linked by a flexible helix. Each of the two domains comprise two EF-hands, so that allowing the protein to bind four calcium ions. The sequence homology among the two domains is high (75%), however, the small differences are able to give different biochemical properties to the respective EF-hands. Indeed, the Kd of the C-terminal EF-hands for calcium is ~10^{-6} M, 10-fold stronger than the N-terminal EF-hands (Kd ~10^{-5}). Calcium binding to CaM induces conformational changes that are subsequently involved in the regulation of several cellular processes. These changes cause the exposition of an hydrophobic surface that interact with targets.

However, also the flexible linker takes part in target binding, since it enables a great deal of variability in the relative orientation of the two domains that is critical for the ability of CaM in the interaction with such a large number of targets [43].

The best known mode of interaction of CaM with targets is described by the binding of both domains to a single binding region and hence is called the “wrap-around” mode. This mode of interaction is observed for all the peptides derived from the autoinhibitory domains of myosin light chain kinases and for other CaM-dependent kinases [44,45], and causes the peptide to adopt an helical conformation, becoming enveloped by the two domains. This was initially considered the predominant mechanism for CaM-binding. However, the subsequent characterization of several structures of the protein in complex with targets showing posttranslational modifications, enlarged the CaM-binding repertoire. This confirm the important roles of the conformational flexibility within each domain to give the possibility of various different relative position of the two domains in order to adapt to different targets. Indeed, CaM is able to bind targets also in an extended mode, so that their domains interact with different regions of the target. The most direct example of this binding mode arise from the structure of the complex of CaM with the anthrax exotoxin, the Edema factor [46]. This mode is also recurrent in the binding of the apo-CaM with targets, and many of these targets (such as neuromodulin and neurogranin) interact with the IQ motif of the protein, which contain the consensus sequence IQxxxRGxxxR [47]. The binding can occur both in presence and in absence of calcium, and, in some cases, through not only the IQ motif by itself, but also combined with other CaM binding sequences. Finally, several examples of CaM-induced dimerization of the target have been reported. An example of this binding mode are present in

the structure of the CaM complex with the gating domain of the Ca^{2+} activated K^{+} membrane channel [48].

1.3.2 Receptor for advanced glycation endproducts

The Receptor for Advanced Glycation Endproducts (RAGE) is a member of the immunoglobulin protein family of cell surface molecules, whose gene is localized on chromosome 6 in the Class III region of the major histocompatability complex[49-52].

The receptor is composed of an extracellular part, a transmembrane helix and a short cytosolic tail. The extracellular region comprise three immunoglobulin-like domains: one variable V-type and two constant C-type, usually referred as C1 and C2[53] (Figure 2).

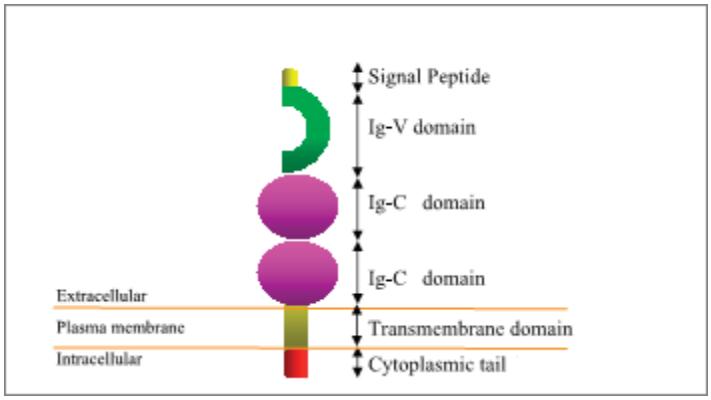


Fig. 2: Schematic representation of RAGE. (Reprinted from Hudson BI et al. FASEB J. 2008)

RAGE is expressed as both full-length (flRAGE), membrane-bound (mRAGE) and various soluble forms. The latter are produced by both proteolytic cleavage of fl-RAGE and alternative mRNA splicing, thus originating isoforms that consist of the extracellular domain lacking the transmembrane and cytoplasmic domains[54,55]. Among the over 20 different splice variants that have been identified to date in humans, the so-called "Endogenous Secretory RAGE" (esRAGE) is the only one that have been detected at the protein level in vivo and is present in a wide variety of human tissues[56].

Although RAGE expression is high in embryonic cells, in a wide range of the differentiated adult cells is expressed at low levels, such as neurons, smooth muscle cells, mesangial cells, mononuclear phagocytes, hepatocytes and cardiac myocytes. An exception is

the lung tissue, in which RAGE is expressed at substantially higher levels[57]. RAGE expression augment in association with inflammation-related pathologies such as vascular disease, cancer, neurodegeneration and diabetes[58].

Binding of ligands to RAGE has been shown to activate multiple cellular signaling cascades, among which ERK1/2-MAP kinases, SAPK/JNK-MAP kinases, JAK/STAT pathway. Many of these signaling cascades result in the activation of the downstream effector NF-κB[59] (Figure 3).

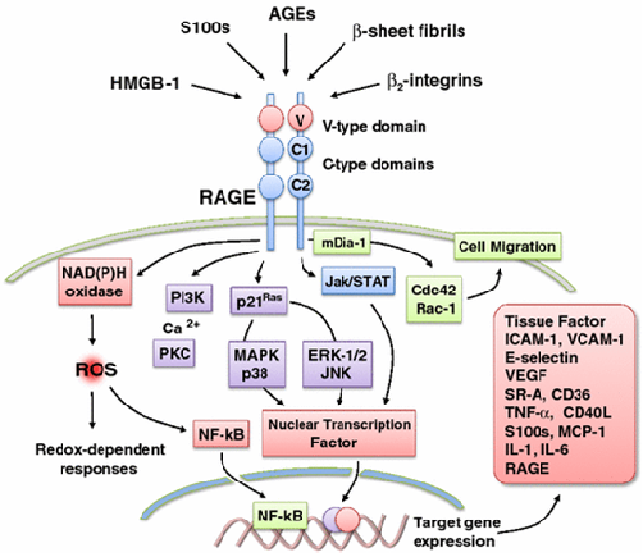


Fig.3: RAGE signal transduction pathways. (Reprinted from Vazzana N. et al. Intern Emerg Med 2009)

RAGE ligands can be grouped into distinct families. The name of the receptor itself account for the first class of molecules that was recognized as RAGE ligand: Advanced glycation endproducts (AGEs) that comprise a class of products of reactions between proteins or lipids and aldose sugars[60]. Besides AGEs, RAGE was found to bind amyloid forming peptides[61], the DNA binding protein amphoterin[62] and S100 proteins[63].

1.3.3 Matrix metalloproteinases

Matrix metalloproteinases (MMPs) comprise a family of at least 28 secreted or transmembrane enzymes collectively capable of processing and degrading various proteins that constitute the complex structural entity that surround and support the cell: the

extracellular matrix (ECM) [64]. They are involved in a number of pathogenic processes including tumour invasion and metastases, rheumatoid and osteoarthritis, angiogenesis and wound healing. At least 22 members of this family have so far been found in human tissues and they are divided into three groups, with respect to the main activity of the purified enzymes in vitro: the collagenases degrade fibrillar collagens type I, II and III; the gelatinases cleave triple helical type IV collagen and gelatine, the stromelysins are active against laminin, fibronectin, proteoglycans type IV collagen and other collagens with interrupted triple helices, and, finally, the membrane-type MMPs (MT-MMPs) are a subgroup of 5 members [65].

All MMPs are synthesized with a prodomain [66] containing a leader sequence, which targets the protein for secretion. They are secreted as latent proforms, with a few exceptions of furin-processed proteinases, such as MMP-11 or MMP-28. The prodomain of MMPs has an egg-like shape, and contains a well conserved cysteine switch motif of PRCXXPD for maintaining the proMMP latent [67,68]. Generally, the structures of all MMP catalytic domains are quite similar. The shape of the catalytic domain is spherical with a flat active site cleft, which extends horizontally across the domain to bind peptide substrates or inhibitors. The catalytic domain has the zinc-binding motif, HEF/LGHS/ALGLXHS, which coordinates a zinc atom at the active site, and under the zinc, a ALMYP methionine-turn. The latency of the zymogen is maintained through cysteine-switch motif [68], in which the cysteine residue acts as a fourth zinc-binding ligand to maintain the enzyme inactive. In addition to the catalytic zinc, the catalytic domain also contains a structural zinc and two to three calcium ions. A sub-site- or S1'-pocket- or channel-like structure is a binding site for a substrate or inhibitor molecule within the active site, and differs considerably in size and shape among the various MMPs. P1' indicates the residue of a bound substrate molecule. The P1'-S1' interaction mainly determines the affinity of inhibitors, and the cleavage positions of peptide substrates. C-terminal hemopexin or vitronectin-like domains affect substrate or inhibitor binding, membrane activation and some proteolytic activities. The hemopexin domain, very similar in structure among the MMPs, is an ellipsoidal disc, and is connected to the catalytic domain by a hinge region. The hinge region is flexible and rich in proline residues [69]. It may also influence substrate specificity.

Hierarchical regulation of MMP activity occurs on many levels [70], including gene expression control, proteolytic activation of MMP zymogens [68], inhibition by endogenous tissue inhibitors of metalloproteinases (TIMPs) [71], and both positive and negative proteolytic feedback loops [72].

Since many pathologies may benefit from control of the activity of MMPs, the quest for suitable human MMP inhibitors (MMPIs) has been actively pursued for more than a decade. Many inhibitors endowed with high-affinity, though modest selectivity, based on a variety of molecular scaffolds have been reported [3], and some entered into clinical trials for different indications, mainly in cancer and arthritis [73].

It has been hypothesized that several side-effects observed during the clinical trials with MMPI could be related to the poor specificity and selectivity of the investigated molecules [74]. These include musculoskeletal pain and inflammation which often required the interruption of the therapy. Several attempts to design selective inhibitors have been carried out by using the structure-based strategy.

1.4 Aims and topics of the research

The research carried out during my PhD has been focused on cloning, expression and characterization of metalloproteins and immunoglobulin cell surface proteins and on the analysis of non covalent interactions relevant for these biological systems. In particular selected examples of: i) protein-protein interactions; ii) interactions between different domains of the same protein and
iii) protein-small molecule interactions have been investigated in their energetic and dynamic aspects.

To date is well established that members of the multigenic S100 protein family play regulatory roles within cells through the interaction with several effector proteins, thus regulating enzyme activities, cell growth and differentiation and calcium homeostasis[17,21,75]. Moreover, secretion has been proved for several members of the S100 protein family and growing evidence suggests that most of the secreted S100 proteins exert their activity through a common receptor: the Receptor for Advanced Glycation Endproduct (RAGE) [76]. The molecular mechanisms of the interaction between S100 proteins and RAGE are poorly characterized, however, evidences collected so far suggest that S100 proteins might form sub-groups which bind to different sites on RAGE. Thus, the localization of critical surface residues involved in RAGE/S100 interaction will be helpful to deduce a more general scheme of ligand recognition and binding by RAGE receptor. S100P has been shown to interact with

RAGE in vitro, since it was shown to trigger the activation of NF-Kb through the MAPK pathway in a RAGE dependent manner in pancreatic and colon cancer cells lines[77,78]. Moreover, the structural determinants of the interaction between RAGE and S100P, as in general with all S100s, are still poor characterized. The aim of the work was therefore focused on cloning, expression and characterization of S100P and on different constructs of RAGE. Then the interaction of S100P with the selected domains of RAGE receptor has been investigated by high resolution NMR spectroscopy, in order to identify the RAGE-S100P interaction surface and to obtain a structural model of the complex.

Efforts have also been directed at the characterization of the structural and dynamical features of a uncommon member of the S100 protein family, which structure is still unknown: S100A16.

S100A16 has been successfully expressed and the solution structure of both the apo and calcium(II) states solved by NMR. Mobility studies have been also carried out in order to obtain the starting point for future investigation of S100A16 interaction with possible targets.

Calmodulin (CaM) is one of the most investigated examples of multidomain protein with the domains experiencing flexibility. The calcium-free form of the protein (apoCaM), which is the resting state of CaM in cells, is able to functionally bind a number of protein targets, as do in the calcium-bound state but its dynamics has received less attention than the latter. It was suggested that helices in apoCaM are quite mobile[79] and that the C-terminal domain experience a conformational exchange. Furthermore, the calcium-binding loops are found particularly unstructured in the calcium free state, most likely due to their high flexibility[79,80]. In order to evaluate the role of the weak interactions in interdomain flexibility and to address the open questions regarding the importance of the relative mobility of the domains, samples of CaM where the exchangeable protein protons are replaced by deuterium have been prepared. The relaxometric analysis performed on these CaM samples allowed us to determine the dynamics of the side chains in the apo state of the protein, the reorientation time value and a collective order parameter, which monitors side chain mobility.

One of the strategies aimed to design molecules with high affinity towards pharmaceutical targets consists in tethering with a suitable linker, low affinity fragment for neighbouring binding sites[81-83]. Thus, the affinity of molecules created in this way will be higher with respect to the sum of the affinity of the single fragments. The gain in affinity is

related to the so-called linking effect. To clarify the energetic aspects of tethering and to understand its limits and possibilities, we have investigated the interaction of a well structurally characterized inhibitor and of its deconstructing fragments with MMP-12 by isothermal titration microcalorimetry, X-Ray, NMR and fluorimetry.

Reference List

1. Smith,C.A. & Wood,E.J. Biological molecules. Chapman & Hall, London (1991).

2. Petsko,G.A. & Ringe,D. Protein structure and function. New Science Press, London (2004).

3. Gavin,A.C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-147 (2002).

4. Ho,Y. *et al.* Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **415**, 180-183 (2002).

5. Nooren,I.M.A. & Thornton,J.M. Diversity of protein-protein interactions. *Embo Journal* **22**, 3486-3492 (2003).

6. Cho,S., Park,S.G., Lee,D.H. & Park,B.C. Protein-protein interaction networks: from interactions to networks. *J. Biochem. Mol. Biol.* **37**, 45-52 (2004).

7. Greer,J., Erickson,J.W., Baldwin,J.J. & Varney,M.D. Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.* **37**, 1035-1054 (1994).

8. von Itzstein,M. *et al.* Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **363**, 418-423 (1993).

9. Williams,D.H., O'Brien,D.P. & Bardsley,B. Enthalpy/entropy compensation as a competition between dynamics and bonding: the relevance to melting of crystals and biological aggregates. *J. Am. Chem. Soc.* **123**, 737-738 (2001).

10. Hajduk,P.J., Huth,J.R. & Fesik,S.W. Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* **48**, 2518-2525 (2005).

11. Huth,J.R., Sun,C., Sauer,D.R. & Hajduk,P.J. Utilization of NMR-derived fragment leads in drug design. *Methods Enzymol.* **394**, 549-571 (2005).

12. Henikoff,S. *et al.* Gene families: the taxonomy of protein paralogs and chimeras. *Science* **278**, 609-614 (1997).

13. Nelson,M.R. & Chazin,W.J. Structures of EF-hand Ca(2+)-binding proteins: diversity in the organization, packing and response to Ca2+ binding. *Biometals* **11**, 297-318 (1998).

14. Schafer,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 calcium-binding protein family. *Genomics* **25**, 638-643 (1995).

15. Heizmann,C.W., Fritz,G. & Schafer,B.W. S100 proteins: structure, functions and pathology. *Front Biosci.* **7**, d1356-d1368 (2002).

16. Marenholz,I., Lovering,R.C. & Heizmann,C.W. An update of the S100 nomenclature. *Biochimica et Biophysica Acta-Molecular Cell Research* **1763**, 1282-1283 (2006).

17. Heizmann,C.W., Ackermann,G.E. & Galichet,A. Pathologies involving the S100 proteins and RAGE. *Subcell. Biochem.* **45**, 93-138 (2007).

18. Santamaria-Kisiel,L., Rintala-Dempsey,A.C. & Shaw,G.S. Calcium-dependent and -independent interactions of the S100 protein family. *Biochemical Journal* **396**, 201-214 (2006).

19. Heizmann,C.W. & Fritz,G. Handbook on Metalloproteins. Marcel Dekker, New York (2004).

20. Donato,R. Functional roles of S100 proteins, calcium-binding proteins of the EF-hand type. *Biochim. Biophys. Acta* **1450**, 191-231 (1999).

21. Donato,R. Intracellular and extracellular roles of S100 proteins. *Microsc. Res. Tech.* **60**, 540-551 (2003).

22. Donato,R. S-100 proteins. *Cell Calcium* **7**, 123-145 (1986).

23. Nelson,M.R. & Chazin,W.J. Structures of EF-hand Ca^{2+}-binding proteins: Diversity in the organization, packing and response to Ca^{2+} Binding. *Biometals* **11**, 297-318 (1998).

24. Skelton,N.J., Kordel,J., Akke,M., Forsen,S. & Chazin,W.J. Signal transduction versus buffering activity in Ca(2+)-binding proteins. *Nat. Struct. Biol.* **1**, 239-245 (1994).

25. Gerke,V. & Weber,K. The regulatory chain in the p36-kd substrate complex of viral tyrosine-specific protein kinases is related in sequence to the S-100 protein of glial cells. *EMBO J.* **4**, 2917-2920 (1985).

26. Zimmer,D.B. & Dubuisson,J.G. Identification of an S100 target protein: glycogen phosphorylase. *Cell Calcium* **14**, 323-332 (1993).

27. Ruse,M. *et al.* S100A7, S100A10, and S100A11 are transglutaminase substrates. *Biochemistry* **40**, 3167-3173 (2001).

28. Wilder,P.T., Baldisseri,D.M., Udan,R., Vallely,K.M. & Weber,D.J. Location of the Zn(2+)-binding site on S100B as determined by NMR spectroscopy and site-directed mutagenesis. *Biochemistry* **42**, 13410-13421 (2003).

29. Randazzo,A., Acklin,C., Schafer,B.W., Heizmann,C.W. & Chazin,W.J. Structural insight into human Zn(2+)-bound S100A2 from NMR and homology modeling. *Biochem. Biophys. Res. Commun.* **288**, 462-467 (2001).

30. Brodersen,D.E., Nyborg,J. & Kjeldgaard,M. Zinc-binding site of an S100 protein revealed. Two crystal structures of Ca2+-bound human psoriasin (S100A7) in the Zn2+-loaded and Zn2+-free state. *Biochemistry* **38**, 1695-1704 (1999).

31. Moroz,O.V. *et al.* Both Ca2+ and Zn2+ are essential for S100A12 protein -oligomerization and function. *BMC. Biochem.* **10**, 11 (2009).

32. Hwang,J.J., Park,M.H., Choi,S.Y. & Koh,J.Y. Activation of the Trk signaling pathway by extracellular zinc. Role of metalloproteinases. *J. Biol. Chem.* **280**, 11995-12001 (2005).

33. Yu,W.H. & Fraser,P.E. S100beta interaction with tau is promoted by zinc and inhibited by hyperphosphorylation in Alzheimer's disease. *J. Neurosci.* **21**, 2240-2246 (2001).

34. Rustandi,R.R., Baldisseri,D.M. & Weber,D.J. Structure of the negative regulatory domain of p53 bound to S100B(betabeta). *Nat. Struct. Biol.* **7**, 570-574 (2000).

35. Wu,T., Angus,C.W., Yao,X.L., Logun,C. & Shelhamer,J.H. P11, a unique member of the S100 family of calcium-binding proteins, interacts with and inhibits the activity of the 85-kDa cytosolic phospholipase A2. *J. Biol. Chem.* **272**, 17145-17153 (1997).

36. Treves,S. *et al.* Interaction of S100A1 with the Ca2+ release channel (ryanodine receptor) of skeletal muscle. *Biochemistry* **36**, 11496-11503 (1997).

37. Kerkhoff,C., Klempt,M. & Sorg,C. Novel insights into structure and function of MRP8 (S100A8) and MRP14 (S100A9). *Biochim. Biophys. Acta* **1448**, 200-211 (1998).

38. Newton,R.A. & Hogg,N. The human S100 protein MRP-14 is a novel activator of the beta 2 integrin Mac-1 on neutrophils. *J. Immunol.* **160**, 1427-1435 (1998).

39. Huttunen,H.J. *et al.* Coregulation of neurite outgrowth and cell survival by amphoterin and S100 proteins through receptor for advanced glycation end products (RAGE) activation. *J. Biol. Chem.* **275**, 40096-40105 (2000).

40. Ambartsumian,N. *et al.* The metastasis-associated Mts1(S100A4) protein could act as an angiogenic factor. *Oncogene* **20**, 4685-4695 (2001).

41. Cheung,W.Y. Cyclic 3',5'-nucleotide phosphodiesterase. Demonstration of an activator. *Biochem. Biophys. Res. Commun.* **38**, 533-538 (1970).

42. Kakiuchi,S. & Yamazaki,R. Calcium dependent phosphodiesterase activity and its activating factor (PAF) from brain studies on cyclic 3',5'-nucleotide phosphodiesterase (3). *Biochem. Biophys. Res. Commun.* **41**, 1104-1110 (1970).

43. Barbato,G., Ikura,M., Kay,L.E., Pastor,R.W. & Bax,A. Backbone dynamics of calmodulin studied by ^{15}N relaxation using inverse detected two-dimensional NMR spectroscopy; the central helix is flexible. *Biochemistry* **31** , 5269-5278 (1992).

44. Meador,W.E., Means,A.R. & Quiocho,F.A. Modulation of calmodulin plasticity in molecular recognition on the basis of x-ray structures. *Science* **262**, 1718-1721 (1993).

45. Osawa,M. *et al.* A novel target recognition revealed by calmodulin in complex with Ca^{2+} calmodulin dependent kinase kinase. *Nat Struct Biol* **6**, 819-826 (1999).

46. Drum,C.L. *et al.* Structural basis for the activation of anthrax adenylyl cyclase exotoxin by calmodulin. *Nature* **415**, 396-402 (2002).

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

48. Schumacher,M.A., Rivard,A.F., Bächinger,H.P. & Adelman,J.P. Structure of the gating domain of a Ca2+-activated K+ channel complexed with Ca2+/calmodulin . *Nature* **410**, 1120-1124 (2001).

49. 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).

50. Clynes,R. *et al.* Receptor for AGE (RAGE): weaving tangled webs within the inflammatory response. *Curr. Mol. Med.* **7**, 743-751 (2007).

51. Herold,K. *et al.* Receptor for advanced glycation end products (RAGE) in a dash to the rescue: inflammatory signals gone awry in the primal response to stress. *J. Leukoc. Biol.* **82**, 204-212 (2007).

52. Bierhaus,A. *et al.* Understanding RAGE, the receptor for advanced glycation end products. *J. Mol. Med.* **83**, 876-886 (2005).

53. Neeper,M. *et al.* Cloning and expression of a cell surface receptor for advanced glycosylation end products of proteins. *J. Biol. Chem.* **267** , 14998-15004 (1992).

54. Raucci,A. *et al.* A soluble form of the receptor for advanced glycation endproducts (RAGE) is produced by proteolytic cleavage of the membrane-bound form by the sheddase a disintegrin and metalloprotease 10 (ADAM10). *FASEB J.* **22**, 3716-3727 (2008).

55. Galichet,A., Weibel,M. & Heizmann,C.W. Calcium-regulated intramembrane proteolysis of the RAGE receptor. *Biochem. Biophys. Res. Commun.* **370**, 1-5 (2008).

56. Cheng,C. *et al.* Expression profiling of endogenous secretory receptor for advanced glycation end products in human organs. *Mod. Pathol.* **18** , 1385-1396 (2005).

57. Brett,J. *et al.* Survey of the distribution of a newly characterized receptor for advanced glycation end products in tissues. *Am. J. Pathol.* **143**, 1699-1712 (1993).

58. Yan,S.F., Ramasamy,R. & Schmidt,A.M. Receptor for AGE (RAGE) and its ligands-cast into leading roles in diabetes and the inflammatory response. *J. Mol. Med.* **87**, 235-247 (2009).

59. Bierhaus,A. *et al.* Diabetes-associated sustained activation of the transcription factor nuclear factor-kappaB. *Diabetes* **50**, 2792-2808 (2001).

60. Ramasamy,R. *et al.* Advanced glycation end products and RAGE: a common thread in aging, diabetes, neurodegeneration, and inflammation. *Glycobiology* **15**, 16R-28R (2005).

61. Yan,S.D. *et al.* RAGE and amyloid-beta peptide neurotoxicity in Alzheimer's disease. *Nature* **382**, 685-691 (1996).

62. Hori,O. *et al.* 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).

63. Hofmann,M.A. *et al.* RAGE mediates a novel proinflammatory axis: a central cell surface receptor for S100/calgranulin polypeptides. *Cell* **97** , 889-901 (1999).

64. Shapiro,S.D. Matrix metalloproteinase degradation of extracellular matrix: biological consequences. *Curr. Opin. Cell Biol.* **10**, 602-608 (1998).

65. Egeblad,M. & Werb,Z. New functions for the matrix metalloproteinases in cancer progression. *Nat. Rev. Cancer* **2**, 161-174 (2002).

66. Sternlicht,M.D. & Werb,Z. How matrix metalloproteinases regulate cell behavior. *Annu. Rev. Cell Dev. Biol.* **17**, 463-516 (2001).

67. Springman,E.B., Angleton,E.L., Birkedal-Hansen,H. & Van Wart,H. Multiple Modes of Activation of Latent Human Fibroblast Collagenase: Evidence for the Role of a Cys^{73} Active-Site Zinc Complex in Latency and a "Cysteine Switch" Mechanism for Activation. *Proc. Natl. Acad. Sci. USA* **87**, 364-368 (1990).

68. Van Wart,H.E. & Birkedal-Hansen,H. The cysteine switch: A principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family. *Proc. Natl. Acad. Sci. USA* **87**, 5578-5582 (1990).

69. Bode,W. A helping hand for collagenases: the haemopexin-like domain. *Structure* **3**, 527-530 (1995).

70. Woessner,J.F.Jr. & Nagase,H. Matrix Metalloproteinases. *J. Biol. Chem.* **274**, 21491-21494 (1999).

71. Gomez,D.E., Alonso,D.F., Yoshiji,H. & Thorgeirsson,U.P. Tissue inhibitors of metalloproteinases: Structure, regulation and biological functions. *European Journal of Cell Biology* **74**, 111-122 (1997).

72. Patterson,B.C. & Sang,Q.X.A. Angiostatin-converting enzyme activities of human matrilysin (MMP-7) and gelatinase B type IV collagenase (MMP-9). *Journal of Biological Chemistry* **272**, 28823-28825 (1997).

73. Beckett,R.P., Davidson,A.H., Drummond,A.H., Huxley,P. & Whittaker,M. Recent advances in matrix metalloproteinase inhibitor research. *Drug Discovery Today* **1**, 16-26 (1996).

74. Coussens,L.M., Fingleton,B. & Matrisian,L.M. Matrix Metalloproteinase Inhibitors and Cancer: Trials and Tribulations. *Science* **295**, 2387-2392 (2002).

75. Zimmer,D.B., Wright,S.P. & Weber,D.J. Molecular mechanisms of S100-target protein interactions. *Microsc. Res. Tech.* **60**, 552-559 (2003).

76. Donato,R. RAGE: A single receptor for several ligands and different cellular responses: The case of certain S100 proteins. *Current Molecular Medicine* **7**, 711-724 (2007).

77. Arumugam,T., Simeone,D.M., Schmidt,A.M. & Logsdon,C.D. S100P stimulates cell proliferation and survival via receptor for activated glycation end products (RAGE). *J. Biol. Chem.* **279**, 5059-5065 (2004).

78. Fuentes,M.K. *et al.* RAGE activation by S100P in colon cancer stimulates growth, migration, and cell signaling pathways. *Dis. Colon Rectum* **50**, 1230-1240 (2007).

79. Zhang,M., Tanaka,T. & Ikura,M. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nat Struct Biol* **2**, 758-767 (1995).

80. Kuboniwa,H. *et al.* Solution structure of calcium-free calmodulin. *Nature Struct. Biol.* **2**, 768-776 (1995).

81. Shuker,S.B., Hajduk,P.J., Meadows,R.P. & Fesik,S.W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531-1534 (1996).

82. Hajduk,P.J. *et al.* Discovery of Potent Nonpeptide Inhibitors of Stromelysin Using SAR by NMR. *J. Am. Chem. Soc.* **119**, 5818-5827 (1997).

83. Olejniczak,E.T. *et al.* Stromelysin inhibitors designed from weakly bound fragments: Effects of linking and cooperativity. *J. Am. Chem. Soc.* **119**, 5828-5832 (1997).

Chapter 2

Methodologies in structural biology

One of the aims of structural genomics is the structure-based drug design, that has been extended to the study of the protein function. This can be faced by an High-Throughput approach (that characterized structural genomics projects), thus screening a huge number of genes but limiting the rate of success for each target, or by a large bibliographic and bioinformatics research (in functional genomics projects) in order to maximize the rate of success, but, thus augmenting the cloning and expression strategies that have to be performed. However, a multidisciplinary approach remains necessary in order to proceed on the screening of parameters that are necessary for good samples preparation.

2.1 Genome browsing

Bioinformatics is essential for deciphering the huge of data generated by high-throughput experimental technologies, and in organizing information obtained from traditional biology. The ultimate goal of bioinformatics is to reveal the biological information hidden in the mass of data and to do this, various databases and software for prediction can be used as tools for searching gene banks, for the analysis of protein sequences and for the prediction of a variety of protein properties.

Examples are GeneBank (http://www.ncbi.nlm.nih.gov/sites/entrez) or Ensembl (http://www.ensembl.org/index.html), from which nucleotidic sequences can be downloaded; Swissprot (http://www.ebi.ac.uk/swissprot/), for obtain information on the aminoacidic sequence, isoforms or biophysical properties and dbSNPs (http://www.ncbi.nlm.nih.gov/projects/SNP/) for information about predicted or validated SNPs.

In order to select the protein construct with the highest probability of giving a soluble and folded protein, various tools can help in the prediction of the target protein properties and domain organization. Such predictions can indentify the presence of transmembrane regions (http://www.cbs.dtu.dk/services/TMHMM-2.0/); intrinsically unstructured regions (http://iupred.enzim.hu/); N-terminal signal peptide (www.cbs.dtu.dk/services/SignalP/)[1]; “rare” codons that are infrequently used by the host chosen for the expression of the recombinant protein (for E.coli they can be predicted at http://nihserver.mbi.ucla.edu/RACC/). Moreover, genome browsing can help to find proteins sharing the same fold and the same consensus sequence within different genomes (www.ncbi.nlm.nih.gov/BLAST). Domain border definitions can be done by multiple

sequences alignments, since domains sequences are more conserved during evolution in respect to linker regions (http://align.genome.jp/), and then, the identification and analysis of protein domain architectures in completely sequenced genomes can be performed using SMART tool (http://smart.embl-heidelberg.de/)[2]. Finally known and predicted protein-protein direct and indirect) associations, can be found in the STRING database (http://string.embl.de/).

2.2 Cloning

In order to get high yield of soluble proteins, many factors have to be taken in consideration, such as the choice of the vector, of the cloning strategy, and of culture conditions. Of course, the knowledge of the protein characteristic will help in this choice, thus increasing the chance of success.

The choice of the expression system is the first step that have to be faced. To date different expression systems are available, among which bacteria is the most attractive due to low cost, high productivity, and rapid use. However, the rational choice of the adequate host for a specific protein of interest remains difficult and need to be done taking into account the different characteristics of the expression system itself, such as the rate of cell growth, the cost, the expression level, extra- or intracellular expression or the possibility of posttranslational modifications[3]. For the expression of a protein of prokaryotic origin, the obvious choice is to use E. coli, but if the protein that have to be expressed is from an eukaryotic source, the method of choice will depend on more factors. To date, alternative hosts, as mammalian, yeast and insect cell, are more accessible and less costly[4,5]. Cell-free protein synthesis has also a great potential for the expression of problematic soluble and membrane proteins expression[6]. However, especially for the characterization that requires high yield of labeled protein, such as NMR, the E.coli expression system remains the most widely used.

The expression system govern the following choice of the expression vector. Many plasmids are currently available for the E. coli system and the different characteristics that typify them can influence the yield of expressed soluble protein. The basic architecture of an E. coli vector is shown in Figure 1 and it contains an antibiotic-resistance gene that can select only the clones that contain the plasmid, an origin of replication and a regulatory gene for plasmid replication and regulation of the copy number, a promoter that initiates the

transcriptions, a multiple cloning site that enables to clone the gene of interest into the vector. In addition also tags and fusion proteins can be present in the vector, generally followed by protease cleavage sites. These elements, together with the promoters are those that mostly affect the yield and solubility of the recombinant protein. Many promoter systems of E. coli can be used as tools for protein expression, but actually only a few of them are commonly used. A ideal promoter is strong, has a low basal expression level, is easy to induce and is independent from the common components of culturing media. In the matter of fusion tags and partners, they offer several advantages in terms of expression, solubility, detection or purification.

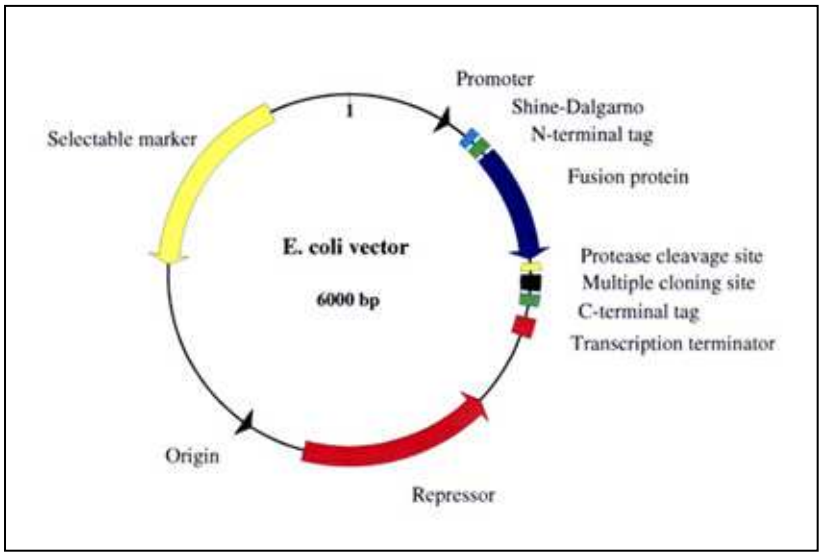


Fig. 1: Basic architecture of an *E. coli* expression vector.

In order to adapt an high-throughput approach to cloning, recently a new technology (Gateway technology) has been developed, thus giving the possibility to clone one or more gene into virtually any expression vector, without the time-consuming reactions that are characteristic of the classic use of restriction enzymes and ligase. This method, described by Landy and co-workers[7] use a site-specific and conservative recombination (LR reaction).

2.3 Site directed mutagenesis

In vitro site-directed mutagenesis is a technique for studying protein structure-function relationships and gene expression, and for carrying out vector modification.

The basic procedure utilizes a vector with the gene of interest and two synthetic oligonucleotide primers, both containing the desired mutation and that are extended during temperature cycling by a high fidelity DNA polymerase. The extension of the oligonucleotide primers generates a mutated plasmid. Following temperature cycling, the product was treated with an endonuclease specific for a methylated DNA, in order to digest the parental DNA template and to select for mutation-containing synthesized DNA (this can be done since DNA isolated from almost all *E. coli* strains is methylated and therefore susceptible to endonuclease digestion). The vector DNA containing the desired mutations is then cloned and expressed by choosing the host and conditions, also in dependence of the single or multiple mutation.

2.4 Protein expression and purification

The screening of different conditions for recombinant protein expression, in order to speed up the production, usually start with a parallel expression of a protein from a variety of vectors containing different tags and/or fusion partners, that are highly expressed and can work as translational enhancer, and a variety of E. coli host strains, that can help by encoding for a number of rare codons, or by reducing proteolytic degradation due to the presence of proteases. The optimization of the expression levels can then be achieved by varying some parameters: the time and/or temperature of induction, that is obtained by the addition of the proper inducer or by changing the growth conditions (in dependence of the kind of promoter used), or the concentration of the inducer itself.

The choice of the strategy for protein purification, depend whether the protein is expressed by the host in the cytoplasm as soluble, or insoluble (in inclusion bodies, IB). Actually the protein can also be forced to be transported in the periplasmic space, due to the presence of a leader sequence (usually pelB and ompT) to the N-terminus of the target protein. If the protein is insoluble, it must be extracted from the inclusion bodies. This is anticipated by extensively wash and centrifuge a several times the IB with buffer containing detergents, such as Triton, or even low concentrations of denaturants, such as guanidine hydrochloride (GdnHCl) of urea. Then the IB are solubilized usually with 4-6 M GdnHCl or 8 M urea.

The purification steps are guided by the physical-chemical and biological characteristics of the proteins: if they are in the native state, ion exchange and size exclusion

chromatography are commonly used, while, when fused with tags, affinity chromatography (Figure 2) is the optimum. As an example, GST-tag is a specific affinity tag, that not only facilitate soluble expression but also increase the efficiency of protein, and can be purified using glutathione immobilized to a matrix such. When the fusion tag is a simple His-tag, purification is achieved by Immobilized metal ion affinity chromatography (IMAC) (that purify proteins with exposed histidine groups via metal ion complex formation).

In order to cleave the tag, various proteases can be used (TEV, Factor Xa, Thrombin, Prescission Protease, recombinant Enterokinase), depending on the protease specific recognition site selected and cloned in the vector codifying for the protein sequence at the cloning step. Then, the cleaved tag can be easily removed by the purified recombinant protein. If the tagged protein is expressed as inclusion bodies, the unfolded protein must be refolded prior to proceed with the cleavage, and this step can be problematic, since sometimes fusion tags may interfere with correct protein folding.

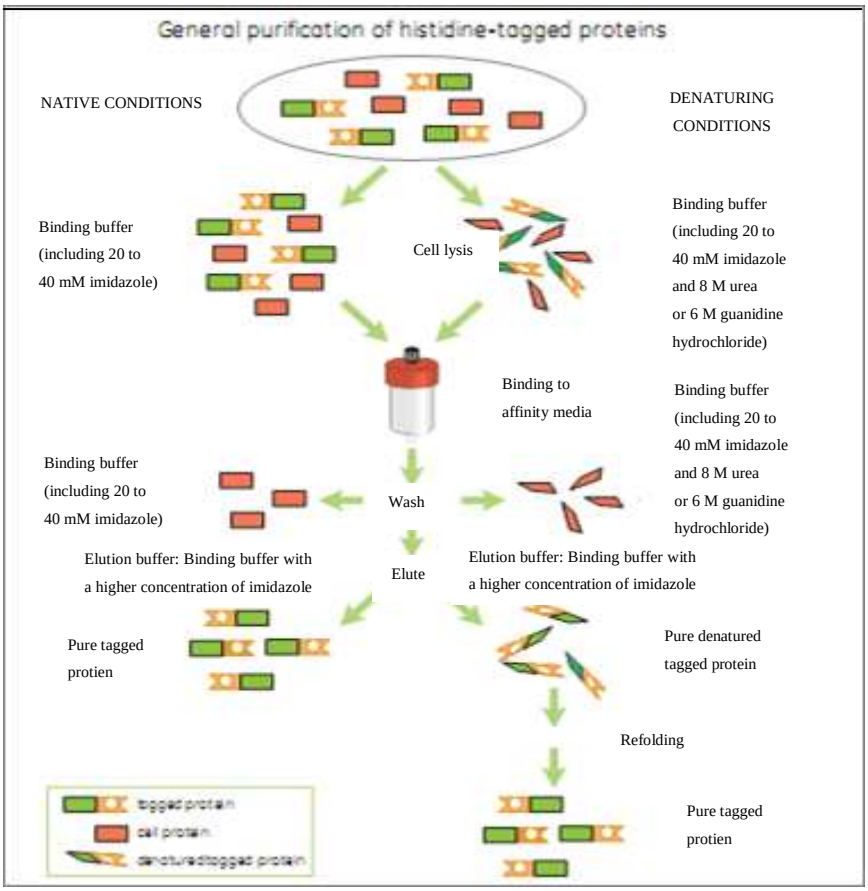


Fig.2: General purification procedure of a typical His-tagged proteins (Reprinted from Recombinant purification handbook for expression and purification of His-tagged proteins, GE Lifescience)

Finally, again structural information can help, or force, in the choice of buffer composition: as example, DTT or Oxidised/Reduced Glutathione (for the presence of disulfide bridges, or reduced cysteine), and EDTA or EGTA (for metal binding proteins).

2.5 Biophysical and structural characterization

2.5.1 Fast Field Cycling Relaxometry

Fast Field Cycling Relaxometry is a NMR technique used for the determination of longitudinal relaxation times of the solvent (T_1) over a wide interval of magnetic fields, ranging from about 10^{-6} to about 1 Tesla. The range boundaries are set by sheer technical issues, being the lower limit affected by the local fields, while the upper limit is mainly determined by technical choices and compromises.

This interval is very wide if compared with the 0.1 T-20T range covered by standard NMR, not considering the impractical technical issues of studying T_1 dispersion curve with an array of standard magnets. On the other hand, FFC relaxometry requires a specialized system, given of much lesser resolution if compared with most NMR spectrometers.

The basic scheme of an experiment can be divided into three phases. At first the sample is polarised in a high field B_p for the time needed to achieve saturation of the nuclear magnetization, then the magnetic field is switched to a lower value B_r for a time t_r, during which the magnetization is allowed to relax towards a new equilibrium value. Eventually, the magnetic field is increased again and the equilibrium magnetization is measured, by applying a 90° pulse followed by acquisition.

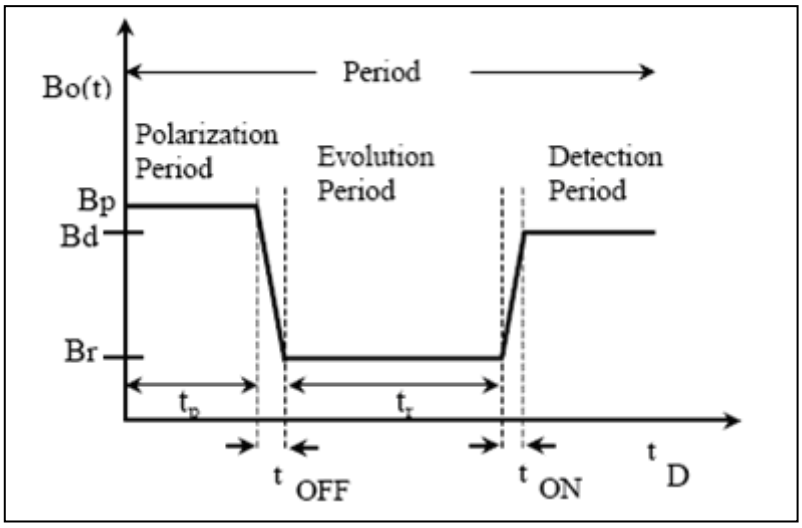


Fig. 3: A typical field cycle in FFCR.

Although FFCR can be applied to several research fields, spanning from dynamics of liquid polymers to proton quantum tunnelling investigation, a useful biological application is the characterization of the hydration of proteins in solution.

When recording the relaxation profile of a protein in D_2O the observed proton relaxation is given by the sum of the contributions of all non exchangeable protein protons.

2.5.2 ITC of protein adduct

Isothermal titration calorimetry (ITC) is a thermodynamic technique that allows the study of the interactions of two species. When these two species interact, heat is either generated or absorbed. By measuring these interaction heats, binding constants (K), reaction stoichiometry (n), and thermodynamic parameters including enthalpy (ΔH) and entropy (ΔS) can be accurately determined. In addition, varying the temperature of the experiment allows the determination of the heat capacity (C_p) for the reaction. The ITC allows researchers to study almost any kind of interaction, including solutes with immobilized enzymes, tissue samples, or other solid materials in suspension. The CSC ITC may also be used to study the decomposition/stability of organic (ex. drugs) and inorganic materials over time (days to weeks). This is particularly useful for determining shelf/storage life of drugs.

An ITC instrument consists of two identical cells composed of a highly efficient thermal conducting material (Hasteloy or gold) surrounded by an adiabatic jacket (Figure 4). The jacket is usually cooled by a circulating water bath. Sensitive thermopile/thermocouple

circuits detect temperature differences between the two cells and between the cells and the jacket. Heaters located on both cells and the jacket are activated when necessary to maintain identical temperatures between all components. In an ITC experiment, the macromolecule solution is placed in the sample cell. The reference cell contains buffer or water minus the macromolecole[8]

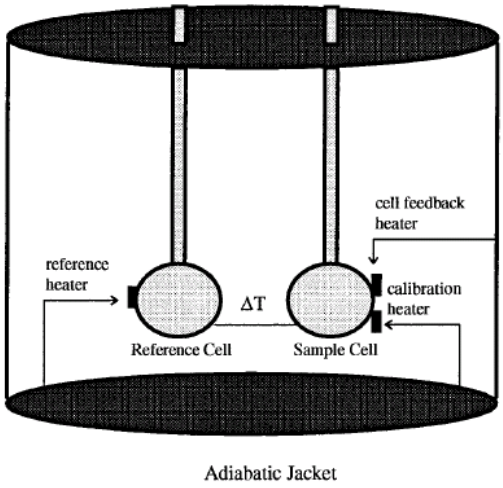


Fig. 4: Schematic diagram of an ITC instrument. Two lollipop shaped cells are contained within an adiabatic jacket[8].

Prior to the injection of the titrant, a constant power[9] is applied to the reference cell. This signal directs the feedback circuit to activate the heater located on the sample cell. This represents the baseline signal. The direct observable measured in an ITC experiment is the time-dependent input of power required to maintain equal temperatures in the sample and reference cell. During the injection of the titrant into the sample cell, heat is taken up or evolved depending on whether the macromolecular association reaction is endothermic or exothermic. For an exothermic reaction, the temperature in the sample cell will increase, and the feedback power will be deactivated to maintain equal temperatures between the two cells. For endothermic reactions, the reverse will occur, meaning the feedback circuit will increase power to the sample cell to maintain the temperature.

The heat absorbed or evolved during a calorimetric titration is proportional to the fraction of bound ligand. Thus, it is of extreme importance to determine accurately the initial concentrations of both the macromolecule and the ligand. For the initial injections, all or most of the added ligand is bound to the macromolecule, resulting in large endothermic or

exothermic signals depending on the nature of the association. As the ligand concentration increases, the macromolecule becomes saturated and subsequently less heat is evolved or absorbed on further addition of titrant. The amount of heat evolved on addition of ligand can be represented by the equation[10]:

$$Q = V_0 \Delta H_b [M]t \, Ka[L]/(1+ Ka[L]) \qquad \textbf{Eq 1.}$$

where V_0 is the volume of the cell, ΔH_b is the enthalpy of binding per mole of ligand, [M]t is the total macromolecule concentration including bound and free fractions, Ka is the binding constant, and [L] is the free ligand concentration.

To determine accurately the enthalpy of binding, it is critical that the first several shots define a baseline region where all added ligand is bound to the macromolecule. The equivalence region should also be well defined by the concentration range spanned by the injections, to determine an accurate value of the association constant. It is necessary that concentrations be chosen so that measurable amounts of free and bound ligand are in equilibrium within the titration zone defined by the titrant injections.

The observed binding isotherm is usually normalized as kilocalories per mole of ligand injected and plotted versus the molar ratio of ligand to macromolecule. The observed heats of binding include contributions from the dilution of the titrant (ligand) and dilution of the macromolecule. The method of data analysis depends on the system of interest. We decided to adopt the procedure for fitting data to the multiple independent binding site model using the analysis software ORIGIN (Microcal, Northhampton, MA) provided with the Omega ITC.

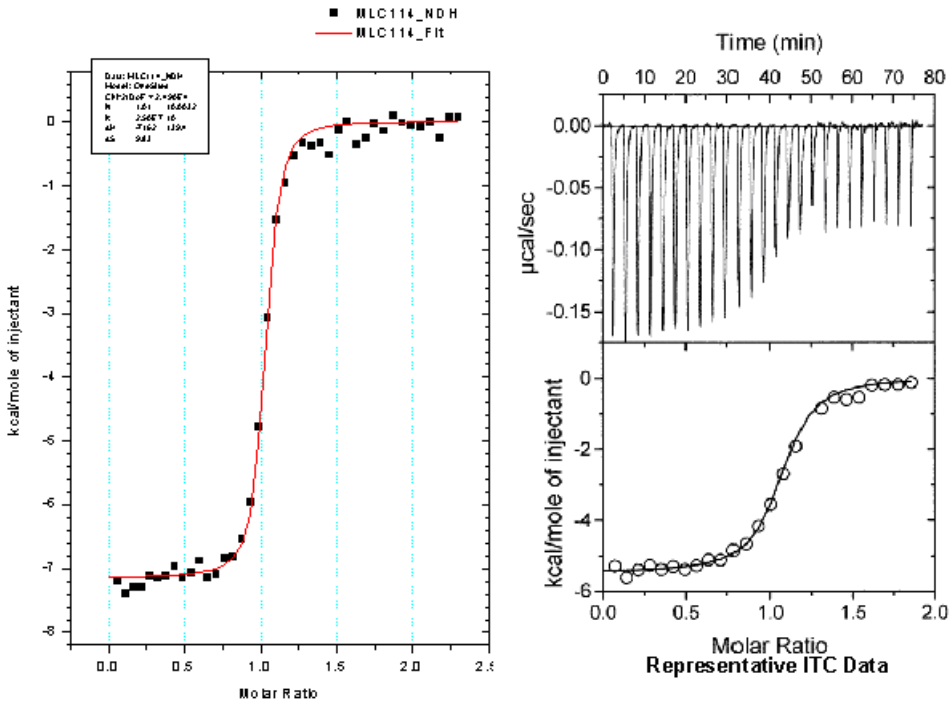


Fig. 5: a) Calorimetric titration of ligand MLC114 with protein catMMP12. The experiment consisted of 41 injections of 8 ml each of a 196mM stock solution of MLC114. b) Representative ITC Data.

2.5.3 NMR spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is unique among the methods available for three-dimensional structure determination of proteins at atomic resolution, since the NMR data can be recorded in solution. Considering that body fluids such as blood, stomach liquid and saliva are protein solutions where these molecules perform their physiological functions, knowledge of the molecular structures in solution is highly relevant. In the NMR experiments, solution conditions such as temperature, pH and salt concentration can be adjusted so as to closely mimic a given physiological fluid. Conversely, the solutions may also be changed to quite extreme non physiological conditions, for example, for studies of protein denaturation.

Furthermore, in addition to protein structure determination, NMR applications include investigations of dynamic features of the molecular structures, as well as studies of structural, thermodynamic and kinetic aspects of interactions between proteins and other solution components, which may either be other proteins or low molecular weight ligands[11,12].

The preparation of the protein sample is a crucial step of this process, since a highly

purified protein sample is required. Inhomogeneous preparation and/or aggregation of the protein as well as low molecular weight impurities may severely harm the structure determination. NMR spectra are then generated by placing the sample in a magnetic field and applying radio-frequency pulses, which perturb the equilibrium nuclear magnetization of those atoms with nuclei of nonzero spin. Transient time domain signals are detected as the system returns to equilibrium. Fourier transformation of the transient signal in to a frequency domain yields a one-dimensional NMR spectrum, which is a series of resonances from the various nuclei at different frequencies, or chemical shifts. The chemical shift of an atom depends on the electronic environment of its nucleus.

NMR spectra of biological macromolecules contain hundreds or even thousands of resonance lines which cannot be resolved in a conventional one-dimensional spectrum (1D). Multidimensional NMR spectra provide both increased resolution and intermolecular correlations which are easy to analyse. The crucial step in increasing the dimensionality of NMR experiments lies in the extension from one to two dimensions. A higher dimensionality experiment consists of a combination of two-dimensional (2D) experiments. All 2D NMR experiments use the same basic scheme which consists of four following, consecutive time periods.

During the excitation period the spins are prepared in the desired state from which the chemical shifts of the individual nuclei are observed during the evolution period t_1. In the mixing period the spins are correlated with each other and the information on the chemical shift of one nucleus ends up on an other nucleus of which the frequency is measured during the detection period t_2. Thus a resonance in the 2D spectrum, a cross peak, represents a pair of nuclei that suitably interact during the mixing time.

Proteins with a molecular weight larger than 10 kDa must be isotope enriched in ^{15}N and ^{13}C for an efficient structure determination; ^{15}N and ^{13}C are used because the most abundant carbon isotope (^{12}C) does not give a NMR signal and the most abundant nitrogen isotope (^{14}N) has undesired NMR properties. The sensitivity obtainable with these types of nuclei greatly varies even if the sample is fully isotope labelled with ^{13}C or ^{15}N. The proton offers the best sensitivity and for this reason constitutes the preferred nucleus for detection of the NMR signal. The other nuclei are usually measured during evolution periods of multidimensional NMR experiments and their information is transferred to protons for detection.

For unlabelled proteins smaller than 10kDa, the combination of the two 2D spectra, [^1H,

^{1}H]- COSY and [^{1}H, ^{1}H]-NOESY often allows the assignment of most proton NMR signals[13]. The first experiment, the [^{1}H, ^{1}H]-COSY (COrrelation SpettroscopY), detects through-bond interactions between protons and correlates protons that are separated by up to three chemical bonds (J couplings). With this experiment the protons within an amino acid can be correlated, however, neighboring amino acids in the polypeptide sequence cannot be connected. The set of correlated proton nuclei is referred to as spin system. However, as soon as an amino acid occurs more than once in a polypeptide chain a direct assignment to a specific sequence position is not possible through COSY experiment. For this purpose the second experiment, the [^{1}H, ^{1}H]-NOESY[14], is measured, where NOESY stands for NOE SpectroscopY[14,15]. The NOE is a consequence of dipole-dipole coupling between different nuclear spin which causes spin polarization to be transferred from one nucleus to any nearby nucleus.; in this way the atoms do not have to be in the same amino acid, they simply have to be close in space[13,14]. The magnitude of the NOE is proportional to r^{-6}, where r is the distance between the interacting nuclei. Unfortunately, the magnitude of the NOE is also affected by a number of other phenomena, such as the rate of tumbling of the protein, which can diminish its magnitude and even make it zero. In practice, NOEs are observed in proteins between hydrogen atoms that are no more than 5 Å apart.

For larger proteins extensive signal overlap prevents complete assignments of all ^{1}H signals in proton spectra. This barrier can be overcome with 3D NMR techniques and uniformly ^{13}C and ^{15}N labelled proteins. The ^{1}H-^{15}N HSQC (Heteronuclear Single Quantum Coherence) is the most important heteronuclear NMR experiment which correlates the nitrogen atom of an NH group with the directly attached proton(s). In ^{13}C, ^{15}N-labelled proteins a sequential assignment strategy can be used which is based on through-bond correlations across the peptide-bond between sequential amino acids. The resolution of NMR spectra can be further increased by including ^{2}H atoms in the protein and by going to the fourth dimension. With these methods systems with molecular weights up to approximately 35 kDa can be studied. Recent advances in both hardware and experimental design promise to allow the study of much larger proteins[16].

When all, or almost all, the resonances of the NMR spectra are assigned, the H-H distances and the dihedral angles, respectively obtained from J^{3} couplings and NOE distances, are used to infer the conformation of the protein. The available programs for the calculation of three dimensional structures utilize, together with experimental constraints, information about the covalent structure of the protein such as the amino acid sequence, bond lengths, bond

angles, chiralities, and planar groups as well as steric repulsion between non-bonded atom pairs. Calculation programs fold a random generated 3D structure, in order to maximize the agreement between the structure and the structural constrains a folded conformation can be determined in great detail. However it must be kept in mind that the experimental constraints do not uniquely describe one exact 3D structure because NMR-derived constraints typically describe a range of possible values and many distances cannot be determined. Thus the result of NMR structure determination is not one model, but a set of similar models, all of which fit the experimentally determined constraints. The RMSD (root mean square deviation) between these models is used to assess how well the structure calculations have converged. Typical structures have backbone RMSD values of less than 1 Å, provided there are not large motions of the backbone or substantially different conformations coexisting in solution.

2.5.4 X-Ray and protein crystallization

X-ray crystallography is essentially a form of very high resolution microscopy. It enables to visualize protein structures at the atomic level and enhances our understanding of protein function. Specifically it is possible to study how proteins interact with other molecules, how they undergo conformational changes, and how they perform catalysis in the case of enzymes. In all forms of microscopy, the amount of detail, or the resolution is limited by the wavelength of the electro-magnetic radiation used. With light microscopy, where the shortest wavelength is about 300 nm, one can see individual cells and sub-cellular organelles. With electron microscopy, where the wavelength may be below 10 nm, one can see detailed cellular architecture and the shapes of large protein molecules. In order to see proteins at atomic detail, we need to work with electro-magnetic radiation with a wavelength of around 0.1 nm or 1 Å. In other words, we need to use X-rays[17].

The diffraction from a single molecule would be too weak to be measurable. So it is necessary to use an ordered three-dimensional array of molecules, in other words a crystal, to magnify the signal. Even a small protein crystal might contain a billion molecules. If the internal order of the crystal is poor, then the X-rays will not be diffracted to high angles or high resolution and the data will not yield a detailed structure. If the crystal is well ordered, then diffraction will be measurable at high angles or high resolution and a detailed structure should result. The X-rays are diffracted by the electrons in the structure and consequently the result of an X-ray experiment is a 3-dimensional map showing the distribution of electrons in the structure.

The data collected from a diffraction experiment is a reciprocal space representation of the crystal lattice[18]. The position of each diffraction 'spot' is governed by the size and shape of the unit cell, and the inherent symmetry within the crystal. The intensity of each diffraction 'spot' is recorded, and is proportional to the square of the *structure factor* amplitude. The *structure factor* is a complex number containing information relating to both the amplitude and phase of a wave. In order to obtain an interpretable *electron density map*, we must first obtain phase estimates (An electron density map allows to build a starting model of our molecule) This is known as the phase problem can be accomplished in a variety of ways.

- Molecular replacement - if a structure exists of a related protein, we can use this structure as a search model and use molecular replacement to determine the orientation and position of our molecules within the unit cell. The phases obtained this way can be used to generate *electron density maps*.
- Heavy atom methods - If we can soak high-molecular weight atoms (not usually found in proteins) into our crystal we can use direct methods or Patterson-space methods to determine their location and use them to obtain initial phases.
- *Ab Initio* phasing - if we have high resolution data (better than 1.6 angstrom or 160 picometers) we can use direct methods to obtain phase information.

Having obtained initial phases we can build an initial model (our hypothesis) and then refine the Cartesian coordinates of atoms and their respective B-factors (relating to the thermal motion of the atom) to best fit the observed diffraction data. This generates a new (and hopefully more accurate) set of phases and a new electron density map is generated. The model is then revised and updated by the crystallographer and a further round of refinement is carried out. This continues until the correlation between the diffraction data and the model is maximized[19].

The bottleneck of X-ray structure determination of macromolecular complex is the crystallization protocol.

In order to crystallize a protein, the purified protein undergoes slow precipitation from an aqueous solution. As a result, individual protein molecules align themselves in a repeating series of "unit cells" by adopting a consistent orientation. The crystalline “lattice” that forms is held together by non-covalent interactions. The non-covalent bonds that hold together the lattice must often be formed through several layers of solvent molecules. In addition to

overcoming the inherent fragility of protein crystals, the successful production of x-ray worthy crystals is dependent upon a number of environmental factors because so much variation exists among proteins, with each individual requiring unique conditions for successful crystallization. Therefore, attempting to crystallize a protein without a proven protocol can be very tedious. Some factors that require consideration are protein purity, pH, protein concentration of protein, temperature, and precipitants. In order for sufficient homogeneity, the protein should usually be at least 97% pure. pH conditions are also very important, as different pHs can result in different packing orientations[19]. Buffers, such as Tris-HCl, are often necessary for the maintenance of a particular pH. Precipitants, such as ammonium sulfate or polyethylene glycol, are compounds that cause the protein to precipitate out of solution.

Two of the most commonly used methods for protein crystallization fall under the category of vapor diffusion. These are known as the **hanging drop**[20] and **sitting drop** methods. Both entail a droplet containing purified protein, buffer, and precipitant being allowed to equilibrate with a larger reservoir containing similar buffers and precipitants in higher concentrations. Initially, the droplet of protein solution contains an insufficient concentration of precipitant for crystallization, but as water vaporizes from the drop and transfers to the reservoir, the precipitant concentration increases to a level optimal for crystallization. Since the system is in equilibrium, these optimum conditions are maintained until the crystallization is complete (see Figure 6).

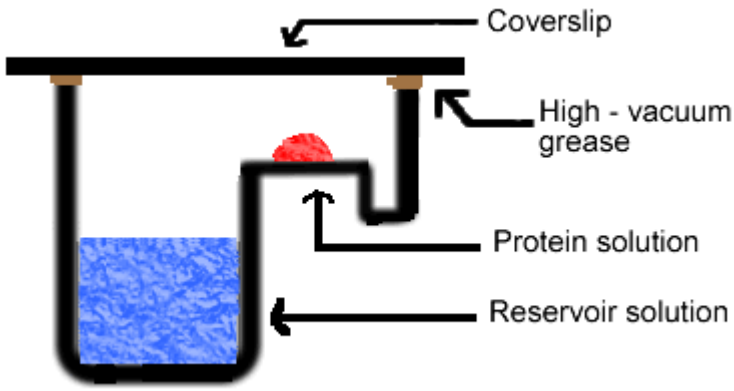


Fig. 6: Solution (blue) usually contains buffer and precipitant. Protein solution (red) contains the same compounds, but in lower concentrations. The protein solution may also contain trace metals or ions necessary for precipitation of particular proteins.

In the past few years macromolecular crystallography has become a standard technique used by many pharmaceutical and biotechnology companies. This methodology offers details of protein-ligand interactions at levels of resolution virtually unmatched by any other technique, and this approach holds the promise of novel, more effective, safer and cheaper drugs. Although crystallography remains a laborious and rather expensive technique, remarkable advances in structure determination and structure based drug design (SBDD) have been made in recent years.

Reference List

1. Emanuelsson,O., Brunak,S., von Heijne,G. & Nielsen,H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**, 953-971 (2007).

2. Letunic,I. *et al.* SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res.* **34**, D257-D260 (2005).

3. Fernandez,J.M. & Hoeffler,J.P. (Accademic Press, San Diego,1999).

4. Boettner,M., Prinz,B., Holz,C., Stahl,U. & Lang,C. High-throughput screening for expression of heterologous proteins in the yeast Pichia pastoris 799. *J. Biotechnol.* **99**, 51-62 (2002).

5. Holz,C., Hesse,O., Bolotina,N., Stahl,U. & Lang,C. A micro-scale process for high-throughput expression of cDNAs in the yeast Saccharomyces cerevisiae. *Protein Expr. Purif.* **25**, 372-378 (2002).

6. Klammt,C. *et al.* Cell-free expression as an emerging technique for the large scale production of integral membrane protein. *Febs Journal* **273**, 4141-4153 (2006).

7. Landy,A. Dynamic, Structural, and Regulatory Aspects of Lambda Site-specific Recombination. Ann.Rev.Biochem. 58, 913-949. 1989. Ref Type: Generic

8. Pierce,M.M., Raman,C.S. & Nall,B.T. Isothermal titration calorimetry of protein-protein interactions. *Methods-A Companion to Methods in Enzymology* **19**, 213-221 (1999).

9. Boynton A.L., MacManus J.P & Whitfield J.F. Stimulation of liver cell DNA synthesis by oncomodulin, an MW 11 500 calcium-binding protein from hepatoma. *Exp Cell Res.* **138**, 454-457 (1982).

10. Indyk,L. & Fisher,H.F. Theoretical aspects of isothermal titration calorimetry. *Energetics of Biological Macromolecules, Pt B* **295**, 350-364 (1998).

11. Wüthrich,K. *et al.* NMR studies of the hydration of biological macromolecules. *Faraday Discuss.* 245-253 (1996).

12. Dyson,H.J. & Wright,P.E. Insights into protein folding from NMR. *Annu. Rev. Phys. Chem.* **47**, 369-395 (1996).

13. Wider,G., Macura,S., Kumar,A., Ernst,R.R. & Wüthrich,K. Homonuclear Two-Dimensional 1H NMR of Proteins. Experimental Procedures. *J. Magn. Reson.* **56**, 207-234 (1984).

14. Kumar,A., Ernst,R.R. & Wüthrich,K. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95**, 1-6 (1980).

15. Jeener,J., Meier,B.H., Bachmann,P. & Ernst,R.R. Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71**, 4546-4553 (1979).

16. Wuthrich,K. The second decade--into the third millenium. *Nat. Struct. Biol.* **5 Suppl**, 492-495 (1998).

17. Scapin,G. Structural biology and drug discovery. *Current Pharmaceutical Design* **12**, 2087-2097 (2006).

18. Rhodes,C.J. & Jacobs,R.L. Self-Energy Calculations in the Hubbard-Model. *Journal of Physics-Condensed Matter* **5**, 5649-5662 (1993).

19. Branden & Tooze. Introduction to Protein Structure. New York (1999).

20. McRee,D.E. Practical Protein Crystallography. Academic Press Inc., San Diego (1993).

Chapter 3

Structural basis of RAGE receptor activation by S100P

In preparation

3.1 Introduction

3.1.1 S100 proteins as RAGE ligands

The first S100 proteins that have been identified as RAGE ligands were S100A12 and S100B[1]. Then, a large number of S100s have been shown to interact with RAGE in vitro and most of them have shown to trigger RAGE dependent signaling in cell-based assays. Structural informations about S100/RAGE engagement were available only for S100B, S100A6 and S100A12. These data suggest that different S100s bind the receptor in different ways (Figure 1). As example, S100B is reported to bind V-domain of RAGE receptor. Binding studies performed by SPR show that the tetrameric form of S100B binds RAGE with higher affinity than the dimeric form, inducing a dimerization of the receptor[2]. It was also suggested that the first C-domain of the receptor participate in S100B binding. On the contrary, S100A6 appears to interact with the second C-domain of RAGE (C2 domain) [3], even if the details of the binding mode are lacking. In case of S10012 the hexamer, formed by the protein in the extracellular space, binds both V and C1 domains and causes RAGE tetramerization[4]. These differences in binding mode, together with the properties of the various cell types in which S100/RAGE binding is reported, might explain the different effects triggered by the interaction.

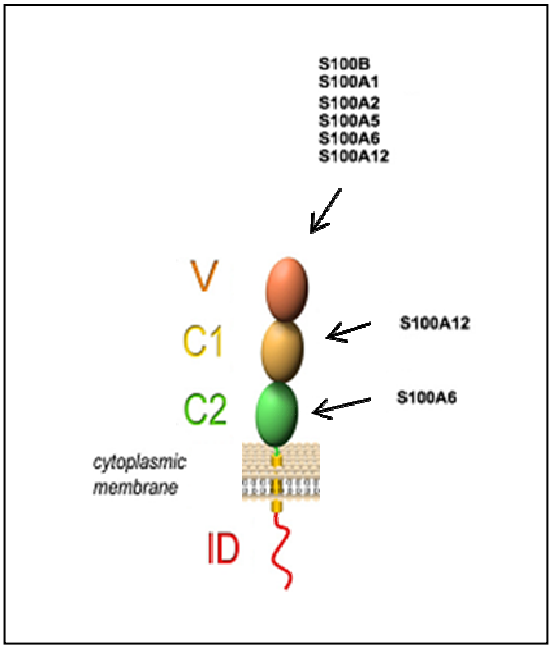


Fig. 1: Interaction of different S100 proteins with different RAGE domains. There are experimental evidence that S100 proteins might form sub-groups which bind to different sites on RAGE. (Reprinted from Leclerc E. et al. Biochimica et Biophysica Acta 2009)

S100P, first purified from placenta (hence its name), is expressed in normal organs or cells and in pathological tissues. This calcium-binding protein plays a role in cytokine-induced differentiation of human myeloid leukemia cells[5] and is present in many tumors including ovarian, pancreatic, gastric, colorectal, breast and prostate carcinomas[6]. It has been suggested that, in tumor tissues, S100P promotes cell proliferation and survival via RAGE engagement through the activation of an ERK1/2-NFκB signaling pathway[7].

The structural determinants of the interaction between RAGE and S100P are still poor characterized. With the aim to clarify the activation of RAGE by S100P, two different RAGE constructs were expressed and tested by NMR for the interaction with S100P. The collected data clearly show that the holo form of S100P interact with the V-domain of RAGE. The alteration of the chemical shifts of both S100P and RAGE monitored on ^{1}H-^{15}N HSQC spectra have been exploited in docking programs to calculate the structure of a possible model of the complex between the two proteins.

3.2 Material and methods

S100P expression. The cDNA encoding for S100P was cloned in pETG30A using the Gateway technology (Invitrogen), in order to obtain a plasmid producing the protein fused with N-terminal His- and GST-tag. The vector was transformed in E.coli strain BL21 GOLD (Novagen) and cells were grown in LB media at 37°C until OD_{600nm} reached 0.6; then protein expression was induced by adding 1 mM IPTG. Cells were allowed to grown at 25°C for 14-16 hours and then harvested by centrifugation at 9000g. Cells were resuspended in 20 mM phosphate pH 7.4, 500 mM NaCl, 10 mM Imidazole and lysis was performed by sonication in ice. The soluble extract, obtained by ultracentrifugation at 40000g, was loaded on a HiTrap chelating HP column (GE Healthcare) previously charged with Ni^{2+} (for selective His-tag binding) and equilibrated with lysis buffer. After washing the column with different step corresponding to different, intermediate Imidazole concentrations (10 mM and 100 mM), recombinant fused protein was eluted with 20 mM phosphate pH 7.4, 500 mM NaCl, 500 mM Imidazole. The protein was then concentrated to 0.7 mg/mL and the cleavage of the tag was performed by AcTev protease assay in 1X TEV Buffer (50 mM Tris-HCl, pH 8.0, 0.5 mM EDTA and 1 mM DTT) at room temperature with overnight incubation. The tag and the AcTev itself (that present a His-tag at the N-terminus), were then removed from the cut

protein by affinity chromatography again with HiTrap column. Purified S100P was then dyalized in 50 mM TRIS pH 7.4, 200 mM KCl, 2 mM $CaCl_2$ (buffer A) in order to perform the final purification step. This was carried out by hydrophobic exchange on Hiprep phenyl FF column (GE Healthcare) equilibrated in buffer A and the protein was eluted with 50 mM TRIS pH 7.4, 200 mM KCl, 5 mM EDTA.

Protein expression and purity was monitored by SDS-PAGE in 17% polyacrylamide stained with Coomasie brilliant blue R-250 against Protein marker.

Cloning and expression of RAGE constructs. Bacterial expression vectors were produced for the single V-domain and the tandem VC1-domain of RAGE. The constructs were amplified by PCR from the flRAGE DNA (GenBank NM_001136) with primers containing 5′ NdeI and 3′ XhoI restriction sites and contained the following RAGE protein sequences (excluding the 22 aminoacid signal peptide): V (23-132) and VC1 (23-243). DNA fragments were subcloned into the pET15b vector (Novagen) in order to express the proteins fused with an N-terminal His_6-tag followed by a thrombin cleavage site. Vectors were transformed in E. coli strain Origami(B) DE3 (Novagen) and cells were grown in LB media at 37°C till an OD_{600} ~ 0.7 was reached. Then expression of the recombinant protein was induced by adding 0.5mM IPTG and growth was allowed for 6 hours at 20°C. Cells were harvested by centrifugation at 9000g and lysed by sonication in ice in 20mM phosphate pH 7.4, 500mM NaCl, 1mM PMSF. Clarified lysate, obtained by centrifugation at 40000g, was first purified on HiTrap chelating HP column (GE Healthcare) equilibrated with the lysis buffer and eluted with a 10 column volume (CV) linear gradient to 20 mM phosphate pH 7.4, 500 mM NaCl, 500 mM Imidazole. Following dialysis in 20 mM phosphate pH 6.0, 150 mM NaCl, the His_6 tag was removed by thrombin cleavage (1 units per mg of protein) incubated at room temperature for 2 h followed by separation over MonoS (GE Healthcare) with a 18 CV linear gradient to 20 mM sodium phosphate, pH 6.0, 850 mM NaCl. Expression and purity of each protein sample was verified by SDS-PAGE in 17% and 15% polyacrylamide stained with Coomasie brilliant blue R-250 against Protein marker.

Expression of isotopically-labelled samples. Samples of ^{15}N- and $^{15}N,^{13}C$-enriched S100P and RAGE constructs were produced as described above except for the use of M9 minimal media containing $^{15}(NH4)_2(SO)_4$ and ^{13}C-glucose as the sole nitrogen and carbon source respectively. $^{2}H,^{15}N,^{13}C$-S100P (~66% as estimated by NMR spectrometry) was

produced in a similar manner through adaptation of E.coli cells to deuteriated conditions achieved in each case by a series of culture enrichments on media with progressively increasing deuterium content.

Expression and purification of E3M-M8Q-M10Q S100P. The mutant E3M-M8Q-M10Q of S100P was obtained using QuickChange site-directed mutagenesis kit (Stratagene) on pETG30A-S100P expressing the WT S100P protein. The company protocol was followed. Several clones coming from each mutation reaction were sequenced. Using the same conditions for the expression of the protein in E. coli cells, the mutated protein precipitated in the inclusion bodies. Thus the procedure to obtain the mutants of S100P was similar to the one used for the wild type protein, the only differences were the addition of 8 M urea to the lysis buffer to solubilize the inclusion bodies, and, after the first purification step carried out by HisTrap chelating FF column, the protein was refolded with a direct step.

NMR. After purification, protein samples of S100P with concentration ranging from 0.3 to 0.8 mM were prepared by buffer exchange in Centricon cutoff 3000 Da (MILLIPORE) by washing with 10 mM HEPES-NaOH pH 7, 75 mM NaCl. The same procedure was followed for prepare NMR samples of V-domain 0.38 mM. For NMR experiments carried out with the holo-S100P, excess $CaCl_2$ was added to each sample up to a final concentration of 10 mM.

NMR experiments were performed on a Bruker Avance spectrometer, operating at a ^{1}H frequency of 500, 800 and 900 MHz (in dependence of the experiment that was carried out) and equipped with cryoprobes. All NMR data were collected at 298 K and 310 K. Titration experiments of holo-S100P with V-domain, and VC1-domain of RAGE were performed by adding 0.2 mM and 0.26 mM solutions of V-domain, and VC1-domain, respectively, into 0.34 mM and 0.45 mM $^{2}H,^{13}C,^{15}N$-S100P solutions. Titration was also carried out for a solution of $^{13}C,^{15}N$-V-domain 0.3 mM with 2.26 mM S100P in three steps to yield V-domain to S100P molar ratios of 1:1,1:2, and 1:3, respectively.

Assignment was performed by a sets of 3D NMR experiments. Decoupling was done on 13C for 2D Trosy and on 2H for 3D trHNCACB and trHNCA. Classical 3D HNCA, CBCACONH and HNCACB watergate were also performed at 500 MHz. Relaxation measurements T1 and T2 and experiments were acquired at 500 MHz. To study interaction, HSQC and Trosy 2D experiments were performed at the 900 MHz.

Docking Calculation. Docking calculation were performed using the program Haddock. Active residues were selected to drive the docking using chemical shift perturbation and solvent accessibility as constraints. Calculation results were grouped in clusters on the basis of energy and RMSD of the complexes.

3.3 Results

Design and expression of RAGE domain constructs. RAGE constructs (V and VC1) were designed accordingly to the sequence alignment with known immunoglobulin family sequences and to the secondary structure prediction. Ala23 was choose as the N-terminus of the constructs since it is the first amino acid of the native receptor after loss of the signaling peptide.

Frequently, in Ig-like domains, pairs of cysteines form disulfide bonds between β-sheet secondary structures[8]. In both V- and VC1-domains structural disulfide bridges are present. Under physiological conditions, the E. coli cytoplasm is maintained in a reduced state that strongly disfavors the formation of stable disulfide bonds in proteins, this is due to the fact that thioredoxins and glutaredoxins (proteins that can catalyze the formation of disulfide bonds in peptides) are maintained in a reduced state by the action of thioredoxin reductase (trxB) and glutathione, respectively. Glutathione in turn is reduced by glutathione reductase (gor). For this reason the oxidation of cysteine thiols in cytoplasmic proteins is strongly disfavored, whether disruption of the trxB and gor genes encoding the two reductases, allow the formation of disulfide bonds in the E. coli cytoplasm[9]. Origami host strains have mutations in both the thioredoxin reductase (trxB) and glutathione reductase (gor) genes. Therefore this strain is usually exploited to express protein containing structural disulfide bonds. For this reason Origami strain has been selected to express both V- and VC1-domains. Protein expression in Origami strain is usually more sensitive to the expression parameters, such as temperature and IPTG concentration which should be careful optimized.

Backbone Resonance Assignments of holo-S100P and V-domain. The chemical shift is a sensitive probe to monitor protein-protein interactions. However the assignment of the protein resonances is essential to identify the regions of the protein involved in the interaction. The assignment and the solution structure of apo-S100P are available on BMRB and PDB

database while the holo form of the protein was not characterized by NMR. The addition of the calcium to the apo protein largely affects the $^{1}H,^{15}N$ HSQC spectra causing an extensive signal broadening. The quality of the spectra was improved by increasing the temperature to 310K and decreasing the protein concentration to 0.5 mM. In order to improve the quality of the spectra and to speed up backbone assignments of holo-S100P, uniformly labeled ^{2}H, ^{15}N, ^{13}C-protein was overexpressed in E. coli. Deuteration reduces the relaxation rates of NMR-active nuclei, and it improves the resolution and sensitivity of NMR experiment. However, incorporation of ^{2}H reduces growth rate of the organism and decrease the protein production as a consequence of the isotopic effect. Thus, bacterial colonies were selected to obtain which better survived during adaptation in high level of $^{2}H_2O$ concentration. E. coli colonies were adapted by repeated subculturing and by a gradual increasing of the deuterium content in the medium. All cell cultures were initiated by inoculation with a sample of the previous culture to an OD_{600} of 0.1 and were grown to stationary phase. Cultures were grown successively on: (i) LB medium/H_2O, (ii) LB medium/30% $^{2}H_20$, (iii) LB medium/60% $^{2}H_20$, (iv) LB medium/90% $^{2}H_20$. Then, the culture was switched to minimal media (enriched in $^{15}N,^{13}C$)/90% $^{2}H_20$ and finally the clone with the highest growth rate was used for grown in minimal media/99% $^{2}H_20$. With this protocol, an enrichment in ^{2}H of ~66% was achieved, as estimated with NMR spectrometry.

In order to obtain more detailed information on the structural features of the V-domain-S100P adduct, ^{13}C-^{15}N enriched sample of the V-domain has been expressed and purified. Triple resonance NMR experiments for backbone assignment were performed on $^{2}H,^{13}C,^{15}N$-S100P and on $^{13}C,^{15}N$-V-domain, at 298K on 500 and 800 MHz spectrometers equipped with cryo-probe.

Chemical Shift Changes upon S100P-RAGE interaction. The interaction of apo and holo S100P with V-domain has been monitored by NMR. Aliquots of V-domain were added to ^{15}N enriched samples of the S100P, in order to prove the interaction (Figure 2).

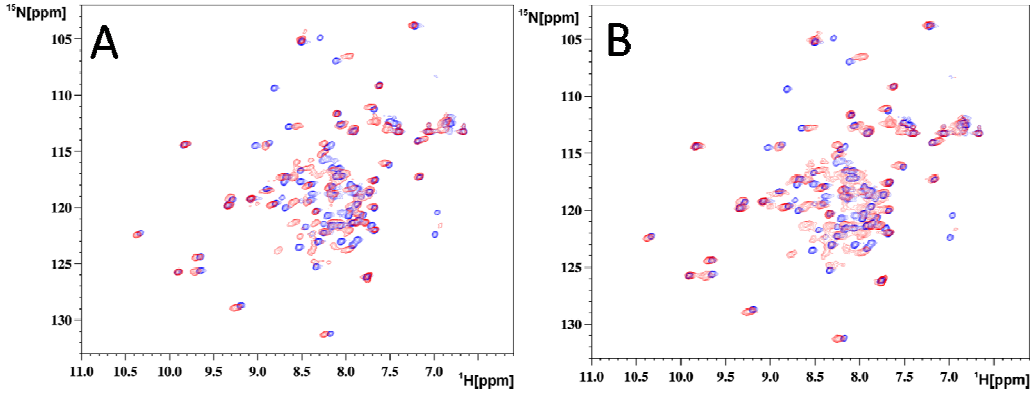


Fig. 2: ^{1}H-^{15}N-HSQC spectrum (800 MHz) of holo-S100P alone (blue), and in presence of V-domain (red) at 310 K (**A**). ^{1}H-^{15}N-HSQC spectrum (800 MHz) of holo-S100P alone (blue), and in presence of VC1-domain (red) at 310 K (**B**).

The evolution of the ^{1}H-^{15}N HSQC spectra clearly show that only the holo-S100P interacts with V-domain, and that the stochiometry ratio is 2:1, i.e. one V-domain binds one S100P dimer. The interaction of holo-S100P with the VC1 construct was also investigated and demonstrated by NMR. The comparison of the two sets of data clearly shows that the presence of the C1-domain does not affect the interaction of the V-domain with S100P (Figure 2).

The effects that originate in the HSQC spectra upon the titration depend on the time scale in which the process is observed. Strong protein-protein interaction are generally in a slow exchange regime on the NMR time scale, thus indicating that the lifetime of the complex is long compared to the change in chemical shift between the free and bound form. This results, in an HSQC experiment, in separate signals that can be observed for the free and bound forms. During the titration, the free form will disappear gradually. On the contrary, in a fast exchange regime, the lifetime of the complex is shorter than the change in chemical shift of the two forms and this is related to weak binding. Only a single averaged signal is observed and this is fractionally weighted according to the populations and chemical shift of the two forms. Finally, in the intermediate condition (intermediate exchange regime), signals will result broadened and arising from both the free and bound forms. Titrating V-domain into holo ^{2}H,^{13}C,^{15}N-S100P several signals became broader, other simply shifted or disappeared. These changes are consistent with the exchange between the free and complexed forms of S100P occurring in the intermediate NMR time scale. Based on these changes, we estimated that the equilibrium-binding constant of V-domain to holo-S100P is in the low micromolar

range. From the point of view of V-domain, upon titration of S100P onto ^{13}C,^{15}N-V-domain, few chemical shift changes were observed with lower intensity comparing with S100P. On the other hand, signal to noise ratio was better and ambiguity lower considering V-domain is a monomer.

Mobility studies on holo-S100P and holo-S100P in complex with V-domain. In order to investigate the structural and dynamical features of the S100P-V-domain complex, relaxation measurements were performed on the holo-^{15}N-S100P and on its complex with the V-domain at 298 K. Calculated values were performed by hydronmr using a model derived from the dimeric X-ray structure where the missing loop (Q46 – K51) was generated by modeler.

The experimental R_1 and R_2 values nicely match with the calculated data obtained using HYDRONMR on the X-Ray structure of the dimeric holo-S100P (PDB 1J55). The residues belonging to the hinge loop (47-52) show slightly higher R1 values, and lower R2 values indicating that this protein region experiences some fast conformational rearrangements. Measurements showed also the six C-terminal residues (E90-K95) present a significant mobility. Again R2/R1 ratio is lower with respect to the rest of the protein. Experimental and calculated data are matching confirming the presence of a dimer molecule (monomer or tetramer would have significantly different values). These data on isolated S100P can be used as a reference for the study of the complex. Relaxation measurement were also performed for ^{15}N-S100P- Vdomain complex at 298K. Results obtained on T1 measurements showed as expected a general decrease of R1 values of S100P residues indicating the formation of a bigger molecule. Interestingly, the mobility observed in the middle loop and at the C-terminal helix is retained. Most probably, these regions are not directly involved in close contact with V-domain but might participate in some conformational change. Calculated data were obtained using hydronmr programme from a Haddock model of the S100Pdimer-Vdomain complex. Again, for the rest of the protein, experimental and calculated R1
data are matching confirming the complex in solution corresponds to a S100P homodimer with V-
domain at the concentration used (0.5 mM for S100P monomer and 0.3 mM for V-domain).

Generation of a model for the S100P/RAGE interaction. Concentration-dependent ^{1}H-^{15}N HSQC experiments performed on the holo-S100P-V-domain complex show intensity attenuation and changes in chemical shift for the residues belonging to the hinge loop region of S100P. These weak effects could hint to the formation of small amounts of S100P tetramer with two V-domains linked to it. Therefore, although present as minor specie in solution, a S100P tetramer bound with two V-domains could be the biologically relevant species on the cell surface. To prove or disprove this hypothesis, models of the possible tetramers have been calculated from the NMR information. The resonances shifted upon the interaction of S100P with V-domain were identified in both the proteins by performing HNCA spectra on the S100P-V-domain complex either enriched in the V-domain or in the S100P partner. Using the chemical shifts perturbations observed on the two proteins, docking calculations were performed by using the program Haddock (Figure 3). The X-Ray structure of the holo S100P and of VC1, and the NMR structure of the V-domain of RAGE receptor (the member of the family closest to the mean PDB) were used as input files for docking calculation. From the analysis of the energetic and scoring functions, it appeared that only one binding mode is meaningful. For both the calculated complex (S100P-V-domain and S100P-VC1 tandem domain) the lowest-energy cluster experienced far much better scoring functions with respect to all others, larger interaction surfaces, and a much lower average number of ambiguous restraints, which indicate better agreement with the experimental data. Looking to the structural features of the two models with the best scoring functions mutations have been planned to abolish the formation of the S100P tetramer.

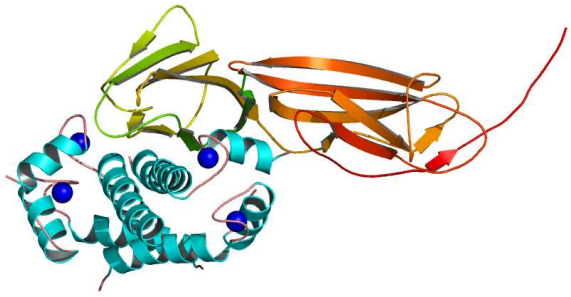
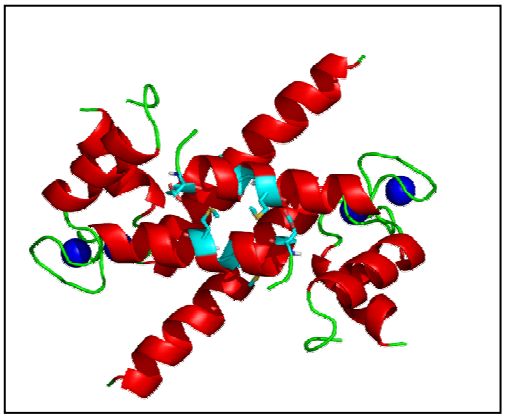


Fig. 3: Ribbod representation of the calculated model of the complex between S100P and VC1-domain.

Production of E3M-M8Q-M10Q mutant of S100P. In the calculated models Glu3, Met8 and Met10 seem to play a crucial role for the stabilization of the S100P tetramer (Figure 4). In this respect the original aminoacids have been replaced with others exhibiting different physical-chemical properties in order to disrupt the interaction responsible for the complex formation.

A



B

Helix E1

```
)P WT:    1 MTELETAMGMIIDVFSRYSGSEGSTQTLTKGELKVLMEKELPGFLQSGKD  50
)P MUT:   1 MTMLETAQGQIIDVFSRYSGSEGSTQTLTKGELKVLMEKELPGFLQSGKD  50

S100P WT:  51 KDAVDKLLKDLDANGDAQVDFSEFIVFVAAITSACHKYFEKAGLK       94
S100P MUT: 51 KDAVDKLLKDLDANGDAQVDFSEFIVFVAAITSACHKYFEKAGLK       94
```

Fig. 4: (A) Ribbon model of the holo-S100P dimer. Aminoacids that are presumably responsible for the stabilization of the S100P tetramer are highlighted in cyana.(B) Sequence comparison of S100P wild type (wt) and the S100P mutant (mut). Changes in the S100P wt sequence to generate the S100P mutant are highlighted in cyana.

The E3M-M8Q-M10Q mutant of S100P was initially expressed and grown in the same, optimized conditions in which the wild type protein was successfully obtained as soluble protein. Unfortunately, the mutant was expressed only in inclusion bodies. Thus, inclusion bodies were first purified from contaminants by washing and centrifuging cells several times in a lysis buffer containing low concentrations of urea. Then, the inclusion bodies containing the protein were solubilized in a 8 M urea buffer The protein was then purified by HisTag affinity chromatography. Then refolding was successfully carried out with a fast 100-fold dilution of urea and then with four steps dialysis against a buffer containing 50 mM TRIS pH 8, 250 mM NaCl. The protein was successively separated from the tag and

purified as already described for the wild type S100P. The proper folding of the protein was verified by 1D ^{1}H-NMR spectra (Figure 5).

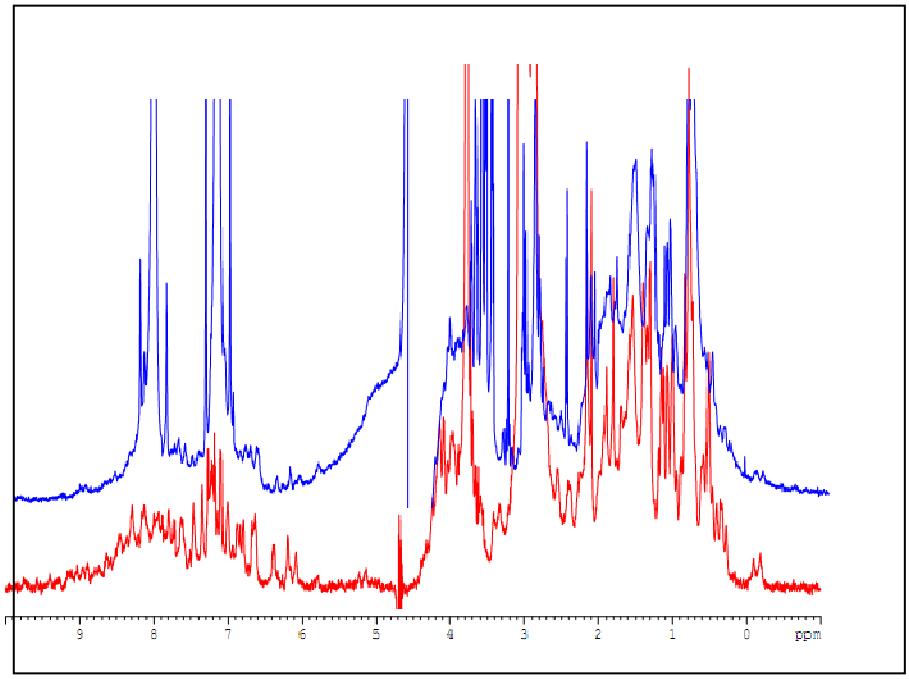


Fig. 5: Superimposition of 1D ^{1}H spectra of S100P wild type (red) and E3M-M8Q-M10Q S100P (blue).

However, the yield of this S100P mutant expressed as His-,GST-tagged construct is very low (~4mg/L in minimal medium). Therefore, two new constructs have been designed and cloned in order to obtain the E3M-M8Q-M10Q mutant of S100P, one fused with only an His-tag at the N- terminus and the other as native protein without tails. The resulting plasmids have been already checked by DNA sequencing and the analysis with different E.coli strains and different growth conditions for a good protein expression performed.

3.4 Discussion

The interaction of S100B with RAGE receptor plays a rule in the observed trophic and toxic effects of the S100 protein in the brain. Moreover, recent results suggest that S100P

contributes to the aggressive nature of pancreatic cancer likely through its ability to activate RAGE. Also in colon cancer the interaction of RAGE with S100P plays an important role as it stimulates cell growth and migration. Therefore, the structural characterization of the RAGE-S100P complex can provide the base to design molecules able to prevent the activation of the receptor. The data collected so far show that RAGE receptor binds the protein with the stochiometry of 1:2. Several signals of the S100P on the HSQC spectra shift upon the addition of increasing concentration of the V-domain of RAGE and some of them enlarge beyond the detection. We have also clarified that only the V-domain is crucial for the interaction with S100P given that V-domain and VC1 tandem construct induce the same pattern of shift on the S100P protein. Considering the concentration of the two proteins at the different steps of the titration and the evolution of the signals in the HSQC, the dissociation constant has been evaluated to be in the low micromolar range, larger with respect to what observed in case of S100B. However should be kept in mind that the high affinity constant found for S100B is related to a binding mode where a tetramer of S100B interacts with two RAGE receptors at the same time, while, in this case, the interaction involve only the domain of a single receptor. The analysis allow us to identify the interaction surface between the V-domain of RAGE and holo S100P so providing the possibility to understand the structural details of the interactions of RAGE receptor with one of its phatologically relevant ligands.

Reference List

1. Hofmann,M.A. *et al.* RAGE mediates a novel proinflammatory axis: a central cell surface receptor for S100/calgranulin polypeptides. *Cell* **97**, 889-901 (1999).

2. Ostendorp,T. *et al.* Structural and functional insights into RAGE activation by multimeric S100B. *EMBO J.* **26**, 3868-3878 (2007).

3. Leclerc,E., Fritz,G., Weibel,M., Heizmann,C.W. & Galichet,A. S100B and S100A6 differentially modulate cell survival by interacting with distinct RAGE (receptor for advanced glycation end products) immunoglobulin domains. *Journal of Biological Chemistry* **282**, 31317-31331 (2007).

4. Xie,J. *et al.* Hexameric calgranulin C (S100A12) binds to the receptor for advanced glycated end products (RAGE) using symmetric hydrophobic target-binding patches. *J. Biol. Chem.* **282**, 4218-4231 (2007).

5. Ishii,Y., Kasukabe,T. & Honma,Y. Immediate up-regulation of the calcium-binding protein S100P and its involvement in the cytokinin-induced differentiation of human myeloid leukemia cells. *Biochim. Biophys. Acta* **1745**, 156-165 (2005).

6. Parkkila,S. *et al.* The calcium-binding protein S100P in normal and malignant human tissues. *BMC. Clin. Pathol.* **8**, 2 (2008).

7. Arumugam,T., Simeone,D.M., Schmidt,A.M. & Logsdon,C.D. S100P stimulates cell proliferation and survival via receptor for activated glycation end products (RAGE). *J. Biol. Chem.* **279**, 5059-5065 (2004).

8. Bork,P., Holm,L. & Sander,C. The immunoglobulin fold. Structural classification, sequence patterns and common core. *J. Mol. Biol.* **242**, 309-320 (1994).

9. Bessette,P.H., Aslund,F., Beckwith,J. & Georgiou,G. Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. *Proc. Natl. Acad. Sci. U. S. A* **96**, 13703-13708 (1999).

Chapter 4

Solution structure and dynamics of S100A16 in the apo and Ca^{2+}-bound states

In preparation

4.1 Introduction

S100A16 is the S100 protein widest distributed in human, and highly conserved in mammals (M&H). Expression of most S100 proteins is actually highly tissue and cell specific, whereas S100A16 expression has been reported in a wide spectrum of human tissues (including brain), analogously to S100A2, S100A13 and S100A14. Up-regulation of S100A16 was found in several cancer tissues, suggesting a function related to malignant transformation or tumor development (M&H). S100A16 expression was up-regulated in tumors of bladder, lung, thyroid gland, pancreas, and ovary. Furthermore, investigation of S100A16 intracellular localization in human glioblastoma cells reveals an accumulation of the protein within nucleoli and a translocation to the cytoplasm in response to calcium stimulation[1].

Among the S100 family, S100A16 is a "particular" member since it presents uncommon characteristics. The N-terminal EF-hand was predicted to be functionally inactive since it comprises 15 aminoacids, and lacks the conserved glutamate residue at the last position, analogously to S100A7. The inability of the N-terminal EF-hand to bind calcium was confirmed by flow dialysis experiments carried out by Heizmann and co-workers[1]. Such experiments (performed in an high ionic strength buffer) revealed one Ca^{2+}-binding site per subunit with K_d of 430 μM, that at physiological conditions would be 2- to 3-fold lower. In such way the affinity of calcium for S100A16 becomes very similar to that of many other S100 proteins. Tryptophan fluorescence variations indicated the occurrence of conformational changes upon calcium(II) binding in the C-terminal EF-hand, that lead to the formation of a hydrophobic patch that could involve the hydrophobic residues in helix 3 and 4 and in the calcium binding loop II. They also showed that S100A16 binds zinc(II) in a different site with respect to calcium(II).

Of at least 25 members found to date in humans, 21 S100 proteins have gene located in the S100A cluster on chromosome 1q21. Exceptions are S100P (located on chromosome 4p16), S100Z (cq14), S100B (21q22) and calbindin D_{9k} (Xp22) [2]. The human chromosomal region 1q21 is structurally conserved during evolution and exhibit several rearrangements which occur during tumor development. The localization of S100A16 gene in the S100A cluster on human chromosome 1q21 (a region in which genomic instability frequently occurs in association with various tumors), together with the finding of up-regulation of this protein in several cancer tissues[3] indicate that S100A16 may have a role in the molecular origin of

certain types of tumors and thus that it deserves structural and functional characterization studies.

Considering the uncommon behavior of S100A16 with respect to calcium-binding, although several S100 protein structures are already available, the structural characterization of S100A16 in

solution was here performed in both the apo and calcium(II) states, to address the exceptional characteristics of this protein within the S100 family. Mobility studies were also performed through relaxation rates analysis. This information represents the starting point for future investigations on the binding with possible targets.

4.2 Material and methods

S100A16 expression. Vector pET21a(+), containing the S100A16 gene and cloned in order to produce protein without tag, was transformed in BL21 Gold E. coli strain (Novagen). Cells were grown in LB medium at 37°C until an optical density at 600 nm was reached., then the protein expression was induced by adding 1 mM IPTG. The culture was allow to grow for 4 hours and then cells were harvested by centrifugation. Cells debris were resuspended in lysis buffer (50mM TRIS pH 8.0, 200mM KCl, 1mM DTT, 0.5mM Pefabloc, 10mM EDTA) and soluble protein was extracted by sonication followed by centrifugation. The cleared lysate was then precipitated by slowly adding streptomycin sulfate to 1% and centrifuging at 15,000g for 20 min. The supernatant was dyalized in 50mM TRIS pH 7.0, 50mM KCl, 1mM DTT, 10mM EDTA (buffer A) and loaded on a Q Sepharose FF (anionic exchange, from Amersham) column equilibrated in buffer A and eluted with a linear gradient to 50mM TRIS pH 7.0, 1M KCl, 1mM DTT, 10mM EDTA. The fractions containing S100A16 were collected, brought to 2mM $CaCl_2$ and dyalized against 50mM TRIS pH 7.4, 200mM KCl, 1mM DTT, 2mM $CaCl_2$ (buffer B). The protein was then purified through hydrophobic exchange with a Hiprep phenyl FF (Amersham) column equilibrated in buffer B and eluted with 50mM TRIS pH 7.4, 200mM KCl, 1mM DTT, 5mM EDTA. A final step of purification was performed with a size exclusion chromatography on a Hi-load superdex 75 16/60 (Amersham) column equilibrated with 20mM MES pH 5.5, 200mM KCl, 1mM DTT, 1mM Pefabloc. Protein expression and purity was checked at every step by SDS-PAGE in 17% polyacrylamide after staining of protein bands with Coomassie Blue R-250 against Protein marker (Novagen).

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

To express the SeMet S100A16 protein, the recombinant expression vector pET21a(+) was transformed into the methionine-auxotrophic E. coli B834(DE3) and growth was carried out with a medium based on a synthetic M9 minimal media (MM) supplemented with glucose, vitamins and amino acids with the exception of L-methionine. Bacterial growth was carried out overnight in the presence of L-methionine (Met) and then the colture was used to inoculate the same synthetic media in the presence of L-selenomethionine (SeMet).

The recombinant SeMet-labelled S100A16 protein was purified and crystallized as for the native protein except that all buffers were degassed and included a reducing reagent to avoid oxidation of selenomethionine, and a chelator to remove traces of metals that could catalyze oxidation.

S100A16 crystallization. The purified protein was concentrated to 14 mg ml^{-1}. Crystallization conditions were established by the sitting-drop vapour-diffusion method. A PEG screening kit (PEG/Ion Screen, Hampton Research) was used for preliminary screening. 2 ul of protein solution were mixed with 2 ul of reservoir solution and equilibrated against 0.8 ml of the reservoir solution.

NMR spectroscopy on S100A16 and structure determination . All NMR experiments for assignment were performed at 298 K on a Bruker 500 MHz spectrometer equipped with cryo-probe. Apo and Ca_2-S100A16 samples were ^{13}C, ^{15}N- labeled, concentrated to 0.3 mM and 0.1 mM for the apo and the calcium-bounded samples respectively and buffered in 20 mM MES pH 5.5, 200 mM KCl, 1 mM DTT, 1mM Pefabloc containing 10% D_2O.

Sequential assignments of the backbone resonance were achieved via 3D HNCO, HNCA, CBCA(CO)NH and HNCACB NMR experiments. Side chain assignments were performed through 3D (h)CCH-TOCSY, HBHA(CO)HN together with ^{13}C-NOESY-HSQC and ^{15}N-NOESY HSQC. Proton-proton distance restraints were derived from the analysis of 2D-NOESY, ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC acquired on Bruker 900 and 800 MHz spectrometers equipped with cryo-probe.

The spectra were processed using TOPSPIN and analyzed with CARA. The structures were calculated using the program CYANA-2.0. The best 30 structures out of the calculated 350 structures of the CYANA family were then subjected to restrained energy minimization

with AMBER 8. 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 was used to evaluate the quality of the structures.

Relaxation measurement. ^{15}N-R_1, R_2, and steady-state heteronuclear ^{1}H–^{15}N NOEs were measured at on a 700 MHz spectrometer using standard pulse sequences, 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. The heteronuclear NOE values were obtained from the ratio of the peak height for ^{1}H-saturated and unsaturated spectra. The heteronuclear NOE values and their errors were estimated by calculating the mean ratio and the standard error from the available data sets. R_1, R_2, and NOE values were obtained for 91 and 96 out of the 102 assigned backbone NH resonances for the apo and the calcium forms, respectively. Estimates of the reorientation time were then calculated with the model-free approach. Theoretical predictions of NH, R_1, and R_2 values for apo-S100A16 and Ca_2-S100A16 were calculated by using the HYDRONMR software program.

4.3 Results and discussion

Protein crystallization. The characterization of S100A16 has been carried out in solid state and in solution. After purification, a single band was obtained by 15% SDS-PAGE, corresponding to a molecular weight of human S100A16 monomer. Then, the PEG/Ion Screen, developed by Hampton Research, was used to provide a rapid screening method for the crystallization of the pure protein sample. The best result was obtained with 8% PEG 3350, 0.2M Di-Sodium tartrate dihydrate, 0.2M Tri-Lithium citrate tetrahydrate, 0.2M Tri-Potassium citrate monohydrate, 0.2M Di-Ammonium hydrogen citrate. Although large hexagonal crystals of S100A16 were obtained, nevertheless the structure of the protein could not be solved, due to the phase problem and to the presence of geminated crystals. In order to overcome this problem, the replacement of methionine residues by seleno-methionine was attempt. The presence of selenium in the protein permitted the collection of multiwavelength anomalous diffraction (MAD) data from a single crystal using synchrotron radiation. For the crystallization of the SeMet-S100A16 were used the same conditions as for the native protein and for checking the SeMet incorporation, MS experiments were performed on both

recombinant proteins, with and without SeMet-incorporation, using MALDI (Figure 1).

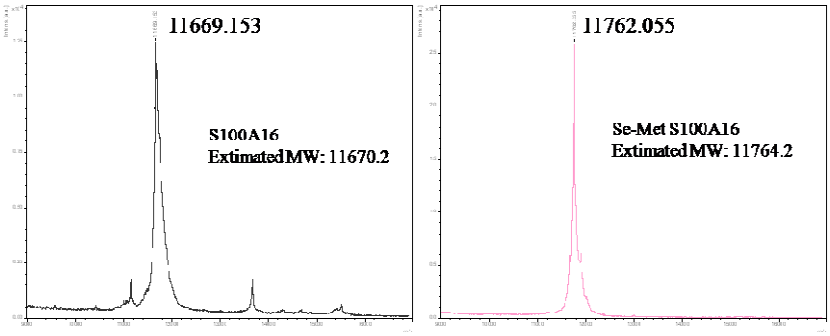


Fig. 1: MALDI mass spectrometry of methionyl-S100A16 (A) and selenomethionyl-S100A16 (B).

The S100A16 expressed by the Met-auxotrophic E. coli B834(DE3)[pET21a(+)/S100A16] in SeMet minimal medium supplemented with L-SeMet was able to incorporate the SeMet amino acid very efficiently. The N-terminal Met of the product was missing, but the next 101 amino acids matched with the known sequence of S100A16 (Accession No. AJ585980). Since the N-terminal Met was removed in vivo, only two Met can be substituted for SeMet in the purified protein. The calculated difference in the molecular mass between the recombinant Met-S100A16 and its SeMet analogue is 93,8. The molecular mass determination by MALDI gave a mass difference of 92,9 (11762 for the SeMet-S100A16 and 11669.1 for Met-S100A16), which matches the expected value very well. Unfortunately the quality of the collected data (Figure 2) was not good enough due to the gemination of crystals.

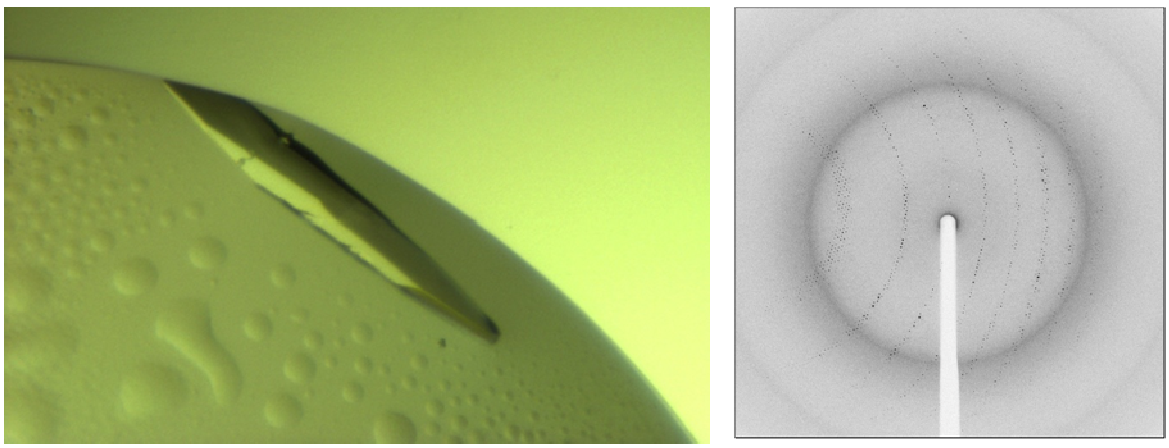


Fig. 2: Crystals and diffraction pattern of purified SeMet-S100A16.

Resonance assignment. The ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) NMR spectra showed well-dispersed signals in both dimensions, which indicated that S100A16 are well folded in both apo and calcium-load state. Besides Ser-1 and Pro-88, the backbone resonance signals of Tyr-19, Lys-31, His-94 in apo-S100A16 and Val-22, Lys-31, Lys-34 in Ca-S100A16 were not assigned.

Ca^{2+} titration of apo-S100A16. The binding of calcium(II) to apo-S100A16 was monitored by following the changes in the ^{1}H-^{15}N-HSQC NMR spectra of ^{15}N-labeled apo-S100A16. Some new peaks appeared in the spectrum during the titration corresponding to the calcium(II)-bound S100A16 form. The intensity of the new peaks increased when the ratio of calcium to S100A16 was increased. When a 1:1 ratio (with respect to the protein subunit concentration) was reached, the original peaks, which correspond to the apo form, disappeared. This kind of behavior is indicative of a slow exchange regime. Figure 3 shows the chemical shift changes on passing from the apo to the calcium(II) form of S100A16. This change (average 0.11 ppm) is much smaller than some other S100 proteins (S100A5 0.5 ppm, S100A13 0.37 ppm).

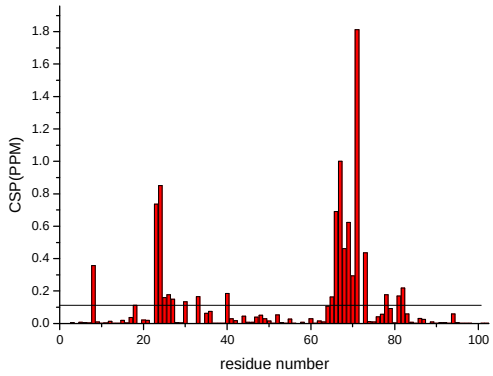


Fig. 3: Composite chemical shift perturbation (CSP), of apo-S100A16 upon calcium(II) binding. The horizontal line indicates the average value.

The residues undergoing the largest changes in chemical shifts are located in the Ca^{2+} binding site, the loop of the C terminal EF-hand motif. These small differences indicate that there are not a significant conformational change occurs like other S100 proteins on passing from the apo form to the calcium form. Flow dialysis experiment yielded S100A16 has a Ca^{2+} binding site with a dissociation constant (K_d) 430 uM[1].

Relaxation measurements. The relaxation parameters for apo and calcium-loaded S100A16 are shown in Figure 4. Such measurements indicate that the protein is dimeric in both forms. The reorientation times corresponding to the observed relaxation rates were calculated to be 12.29±1.46 and 12.34±1.8 ns for the apo and calcium-loaded forms S100A16, respectively, in agreement with the molecular weight and the reorientation times observed for other S100 homodimer proteins[4-8].

apoS100A16 Ca_2S100A16

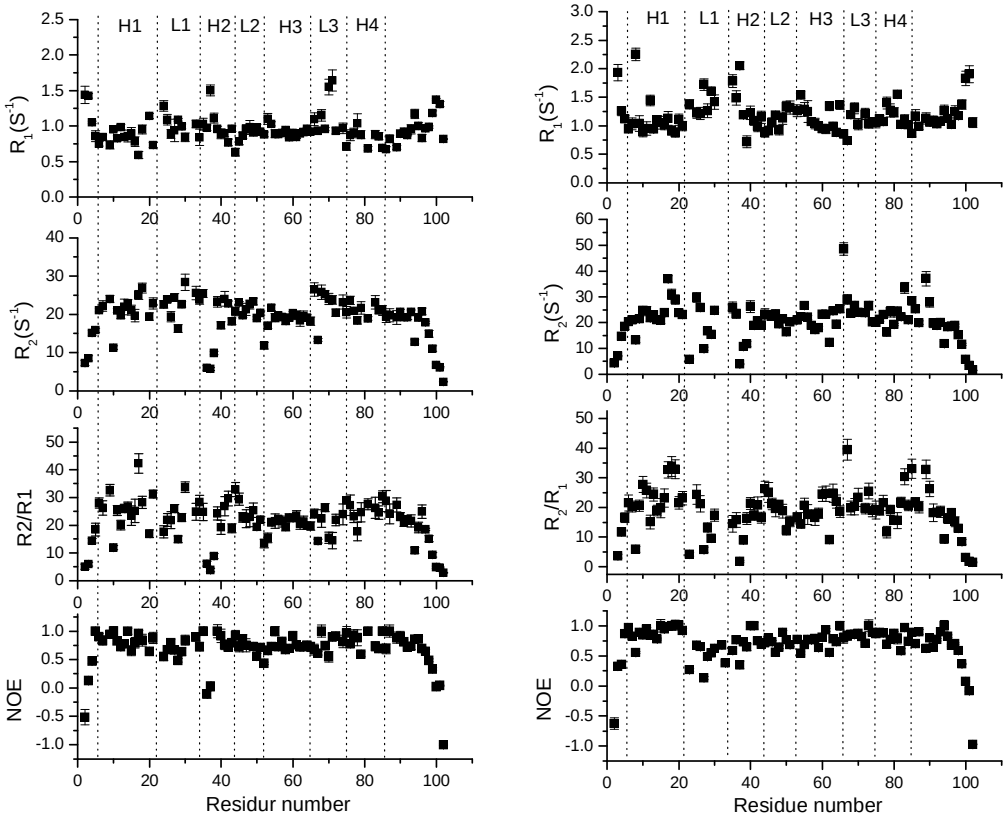


Fig. 4: Sequential plot of the experimental relaxation parameters of apo and Ca^{2+}-S100A16. The values calculated by HYDRONMR are shown as bars. NOE: nuclear Overhauser effect.

In both apo and calcium-loaded S100A16, the residues in the extreme N terminus, the loop of the N terminal EF-hand motif and the C terminus are poorly structured as a result of their fast internal mobility, revealed by the small or negative NOE values. Occurrence of motion is also detected from some residues on the beginning of the α helix II (34, 36, 37 in the apo form; 35, 37 in the calcium form). Upon the calcium binding, the loop between helix I and II has more

motion (the ^{1}H-^{15}N-NOE values are even lower than the apo S100A16) and the residues at the end of the helix IV (83, 85, 89 and 90) and Asp-66 were observed have a motion on a slower time scale, as indicated by the significantly larger R_2 value. This mobility may originate from backbone amide conformational exchange and/or side-chain rotation.

Solution structure of apo and calcium-loaded S100A16. The solution structures of the human S100A16 in the apo and calcium-loaded form were obtained.

Few NOE patterns were detected for the residues in the loop between helix I and II and the C-terminus, consistent with the observed mobility in these regions. In the calcium form, the Ca2+ ion were restrained to be within 3 Å from the oxygen ligand atoms (OD1 of Asp-66, Asn-68; OD1 and OD2 of Asp-70;O of Arg-72;OE1 and OE2 of GLU-77). No symmetry constraint was used. The calculated families of structures are shown in Figure 5.

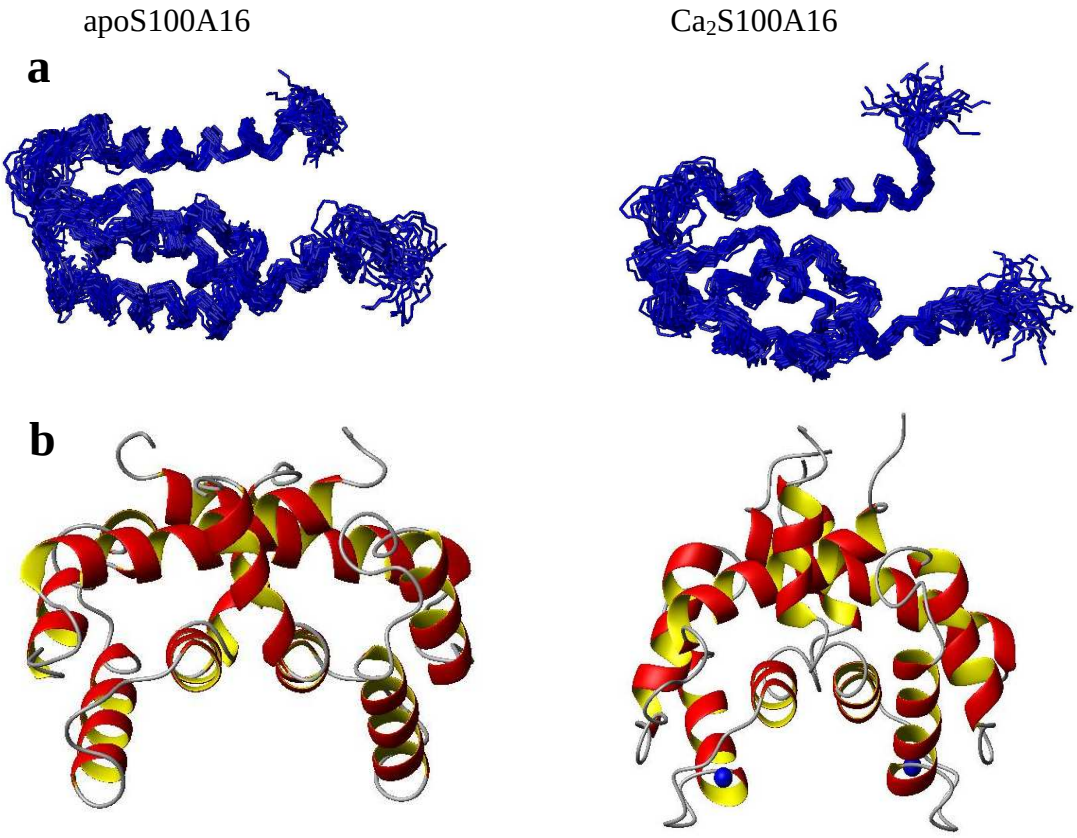


Fig. 5: Families of the solution structures of the S100A16 subunit in the apo and calcium(II)-loaded forms (a) and ribbon representation of the homodimer mean structures (b) obtained after restrained energy minimization.

In both forms, the four helices of the two EF-hand motifs of each subunit present in the dimeric structure are well defined; the loop of the fist EF-hand motif is less well defined; whereas the N and C terminus are poorly defined. These results are in line with the relaxation results.

The root mean square deviation (RMSD) from the mean dimeric structure for the structured regions of the protein is 0.94±0.11 Å (backbone) and 1.33±0.1 Å (heavy atoms) for apo-S100A16 (residues 6-24,34-67,75-94) and 0.83 ±0.11 Å (backbone) and 1.30 (heavy atoms) for Ca_2S100A16 (residues 3-22, 34-94).

In both the apo and calcium-loaded forms, dimerization occurs through interactions between helices I and I' and between helices IV and IV', which form an X-type bundle. At the dimer interface, residues in the hinge between helix II and III make contacts with residues near the N terminus of helix I of the other monomer. The aromatic residues Tyr-78 and Trp-79 in helix IV also make several contacts with helices I' and IV' of the other monomer. In the S100A16 dimer, all these interactions align helix I and IV in opposite directions to helix I' and IV', respectively.

Reference List

1. Sturchler,E., Cox,J.A., Durussel,I., Weibel,M. & Heizmann,C.W. S100A16, a novel calcium-binding protein of the EF-hand superfamily. *J. Biol. Chem.* **281**, 38905-38917 (2006).

2. Ridinger,K., Ilg,E.C., Niggli,F.K., Heizmann,C.W. & Schäfer,B.W. Clustered organization of S100 genes in human and mouse. *Biochim. Biophys. Acta* **1448**, 254-263 (1998).

3. Marenholz,I. & Heizmann,C.W. S100A16, a ubiquitously expressed EF-hand protein which is up-regulated in tumors. *Biochem. Biophys. Res. Commun.* **313**, 237-244 (2004).

4. Inman,K.G., Baldisseri,D.M., Miller,K.E. & Weber,D.J. Backbone Dynamics of the Calcium-Signaling Protein apo-S100B as Determined by 15N NMR Relaxation. *Biochemistry* **40**, 3439-3448 (2001).

5. Zhukov,I., Ejchart,A. & Bierzynski,A. Structural and Motional Changes Induced in apo-S100A1 Protein by the Disulfide Formation between Its Cys 85 Residue and □-mercaptoethanol. *Biochemistry* **47**, 640-650 (2008).

6. Dutta,K., Cox,C.J., Basavappa,R. & Pascal,S.M. 15N Relaxation Studies of Apo-Mts1: A Dynamic S100 Protein. *Biochemistry* **47**, 7637-7647 (2008).

7. Bertini,I., Fragai,M., Luchinat,C. & Parigi,G. ^{1}H NMRD profiles of diamagnetic proteins: a model-free analysis. *Magn. Reson. Chem.* **38**, 543-550 (2000).

8. Arnesano,F. *et al.* Structural interplay between calcium(II) and copper(II) binding to S100A13 protein. *Angew. Chem. Int. Ed Engl.* **44**, 6341-6344 (2005).

Chapter 5

Global and local mobility of apocalmodulin monitored through fast field cycling relaxometry

Borsi, V., Luchinat, C., and Parigi, G.

Biophys J., 97(6), 1765-61, 2009

5.1 Introduction

Calmodulin, a protein which domains experience mobility, has been selected as a model to demonstrate the potential of the collective non-exchangeable protein protons relaxation rate analysis for dynamic characterization of biomolecules. To this aim, samples of calmodulin have been expressed and exchangeable protein protons have been replaced with deuterium.

Calmodulin is one of the most investigated examples of proteins which domains experience mobility [1-7]. The structures of both calcium-bound and apo-CaM have been solved both in the solid state [8,9] and in solution [1,2,10]. The structure of the holo-CaM in the solid state is characterized by an extended conformation; on the other hand, the X-ray structure of apoCaM indicates that the N- and C-terminal domains are close to one another and interact directly [8], thus leading to a globular shape. Moreover, solution NMR studies on both forms indicate high flexibility in the central linker region and the absence of stable, direct contacts between the two domains, which are thus free to reorient and experience multiple relative conformations [2,10].

Among the various experimental techniques developed in order to characterize protein mobility, NMR is known to be a powerful tool [11-18]. Indeed, nuclear spin relaxation measurements (T_1, T_2 and NOE) give information on motions on the picosecond to nanosecond and microsecond to millisecond time scales [6,19,20], $T_{1\rho}$ and spin-echo measurements on the microsecond to millisecond time scale [21], and residual dipolar couplings on the picosecond to millisecond time scale, thus covering also the microsecond to nanosecond range [11,22]. Such measurements are performed on single nuclei of each protein residue, and thus can be used to monitor the presence of motions related to the individual residues. On the other hand, detailed information of the spectral density function may be recovered, but this is not straightforward.

To date, several NMR techniques have been used to characterize in detail the mobility of calcium-bound CaM [5-7,19,23-25], whereas fewer experiments were performed to characterize the dynamics of apoCaM. Backbone amide hydrogen exchange experiments suggested that helices in apo-CaM are quite mobile [10], and the substantial line broadening observed for a number of residues, especially in the C-terminal domain, indicates that the latter experience a conformational exchange. This supports the idea that flexibility is fundamental for the interaction capability of the C-terminal domain of apo-CaM with its target peptides.

Furthermore, the calcium-binding loops are found particularly unstructured in the calcium free state, most likely due to their high flexibility [2,10].

Mobility of both apoCaM and calcium-bound CaM in solution has been studied by measuring the amide relaxation rates, a technique providing information on the mobility of the protein backbone. In this way it was found that the order parameter for all secondary structure elements is of the order expected for well-folded proteins (S^2~0.85 for apoCaM and 0.72 for Ca_4-CaM) when an isotropic diffusion model is used, or measurements are performed on truncated single terminal domains of the protein [6,19,20,26,27]. S^2 values down to ~0.55 are obtained for both forms when a global internal motion model is applied to fit the data [19,20]. However, whereas the relative mobility of the protein domains of calcium-bound CaM has been extensively studied [5,7], it is not clear how much such mobility is important in the apo state of the protein, as the latter has a compact closed structure in the solid state, and such structure may represent the most representative conformation also in solution. Furthermore, whereas in calcium-bound CaM side chains have been shown to be exceptionally mobile [25], no information is available on the mobility of the side chains of CaM in the apo form. In this work, we used high sensitivity fast field cycling (FFC) relaxometry with the aim of addressing the open questions for apo-CaM.

Fast field cycling (FFC) relaxometry is a low resolution technique which allows the measurement of nuclear longitudinal relaxation rates as a function of the magnetic fields, from 0.01 to tens or hundreds of MHz [28,29]. In this way the spectral density function of the observed nuclei (typically the solvent water protons) can be directly accessed, and for this reason it is routinely used e.g. for the study of contrast agents for magnetic resonance imaging [30-38]. Recently, it has been shown that relaxometry can be used to detect the collective relaxation rate of protein protons, thus obtaining direct information on their spectral density function [39-42]. On the other hand, the intrinsically low sensitivity of this technique provides an unique not resolved signal, so that no information can be obtained on the single protein residues. However, these measurements can be used to define a collective order parameter, S_C^2, which reflects the presence of motions in a range of time scales from the picoseconds to the nanoseconds [39,41]. S_C^2 is defined from the reduction in the collective relaxation rate of all non-exchangeable protein protons with respect to the value expected for a rigid protein. Therefore, S_C^2 depends on the motional averaging of all kinds of proton-proton interactions, including long-range interactions, which are certainly more sensitive to internal motions. Even short-range interactions may be heavily influenced by local motions especially for side

chain residues or residues on the protein surface. The measurement of S_C^2 can thus provide a further piece of information for the description of the dynamics of proteins.

5.2 Materials and methods

Purification of apoCaM and sample preparation. Vector pET16b, containing the N60D mutant of CaM gene fused with a His_{10} tag, was transformed in E. coli BL21 GOLD. Cells were grown at 37°C in LB medium to an absorption of 0.8 at 600nm. Then IPTG was added to a final concentration of 0.5 mM, to induce protein expression and the growth was allowed for five hours at 37 °C. Cells were harvested by centrifugation at 9000g and lysis was performed in 20 mM MOPS pH 8, 300 mM NaCl, (buffer A) by sonication in ice followed by centrifugation at 40000g. The soluble fraction containing the His_{10}-tagged protein was loaded on an HisTrap chelating FF column (GE Healthcare) equilibrated with buffer A. After washing the column with an intermediate Imidazole concentration (10mM), recombinant fused protein was eluted with 20 mM MOPS pH 8, 300 mM NaCl, 150mM Imidazole. The protein was loaded on a PD-10 desalting column (GE Healtcare) and the buffer was exchanged to 50mM TRIS pH 8, 100mM NaCl, 1mM $CaCl_2$. Restriction protease Factor Xa (Roche diagnostic) was added to the solution (30 µg of enzyme for 30 mg of recombinant protein) to cut the His_{10}-tag and the reaction was incubated overnight at room temperature. The cleaved tag was separated from the protein by HisTrap chelating FF column and the recombinant protein without the tag was eluted in buffer A. A final step of purification was performed with a size exclusion chromatography on a Hi-load superdex 75 16/60 (Amersham) column equilibrated with 20mM MES pH 6.8, 200mM KCl. The purity of the protein was checked by SDS-PAGE in 15% polyacrylamide gels after staining of protein bands with Coomassie Blue R-250 and the protein yield was about 25 mg/L.

Samples were washed consecutively three times with a 20 mM EDTA solution pH 6.5 to obtain the apo form, then were lyophilized and successively dissolved in D_2O. Protein concentration was about 0.6mM and the pD was adjusted to 7.

Relaxation measurements. Magnetization decay/recovery curves for CaM samples have been obtained at 298 K from 0.02 to 30 MHz using a Stelar fast field cycling relaxometer of the last generation [28]. This technique consists in i) a preparatory part, during which a magnetization is induced through a polarization field (for low frequency

measurements) or a null magnetization is created (for high frequency measurements), ii) a relaxation part, during which the sample is kept at the relaxation field for a series of time intervals, and iii) a detection part, during which a field of 13 MHz is applied and the magnetization arising after such time intervals is measured. The acquisition is performed by applying a 90° radiofrequency pulse and integrating the resulting proton free induction decay (FID). The sensitivity of the instrument has been recently improved so that the signal of protein protons in submillimolar protein solutions in D_2O can be detected with a good signal-to-noise ratio [41]. Magnetization decays from a prepolarized intensity were measured for relaxation fields between 0.02 and 10 MHz, using a polarization time of 0.2 s and a polarization field of 30 MHz. The points in the decay were acquired at 48 times logarithmically scaled between 0.001 to 0.1 s. For the higher fields (10-30 MHz), magnetization recoveries were measured through acquisition of the magnetization signal for 64 times logarithmically scaled between 0.002 to 0.8 s. The number of scans was 512 and 128 for the decay and recovery curves, respectively.

The "universal" distribution of the relaxation rates reported in [41] was used to fit the magnetization decay/recovery. The fit function for the magnetization decay curves (up to 10 MHz) was M=P1+P2*(0.0076*exp(-P3*t*0.11)+ 0.0079*exp(-P3*t*0.13)+ 0.0051*exp(-P3*t*0.16)+ 0.0098*exp(-P3*t*0.19)+ 0.0166*exp(-P3*t*0.229)+ 0.0234*exp(-P3*t*0.275)+ 0.0267*exp(-P3*t*0.331)+ 0.0316*exp(-P3*t*0.398)+ 0.0532*exp(-P3*t*0.479)+ 0.0673*exp(-P3*t*0.575)+ 0.0784*exp(-P3*t*0.692)+ 0.102*exp(-P3*t*0.832)+ 0.1223*exp(-P3*t*1.)+ 0.0881*exp(-P3*t*1.202)+ 0.0566*exp(-P3*t*1.445)+ 0.0588*exp(-P3*t*1.74)+ 0.0751*exp(-P3*t*2.09)+ 0.0849*exp(-P3*t*2.512)+ 0.0575*exp(-P3*t*3.02)+ 0.0208*exp(-P3*t*3.63)+ 0.0069*exp(-P3*t*4.37)), where M is the magnetization, t is the time and P1, P2 and P3 are fit parameters. The relaxation rates present in the distribution are thus provided by P3 times the different numbers in the exponential functions. For the magnetization recovery curves (for frequencies larger than 10 MHz), the same function is used with exp() substituted to 1-exp() and an additional term (1-exp(-0.06*t) is added to include the relaxation recovery of free water protons in D_2O.

5.3 Results

The protocol for the expression and the purification of CaM was developed and then optimized in order to achieve a good yield in soluble protein (25mg/L of colture).

After purification by HisTrap chelating Fast Flow and Hi-load Superdex 75 chromatography, one 16.6 kDa band, corresponding to CaM was visible on reducing 15% SDS-PAGE analysis (Figure 1A). Moreover, the elution profile of the ion-exchange chromatography showed only one peak, confirming the presence of only one form of CaM (Figure 1B).

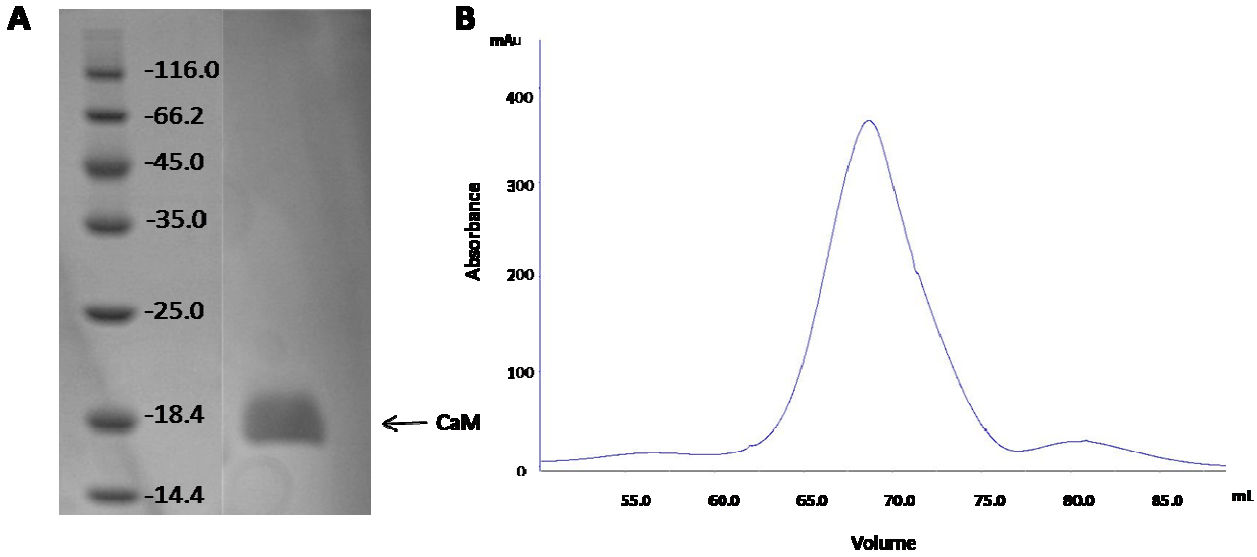


Fig. 1.: (A) Reducing SDS-PAGE of CaM N60D after purification. (B) Elution profile of the protein subjected to ion-exchange chromatography showing only one peak.

CaM was treated with EDTA in order to remove any traces of metal ions from the solution and possible divalent ion bound to the protein. Then the sample was extensively washed with centricon in pure water to remove salt, buffer and EDTA. The sample constituted by CaM and pure water was rapidly frozen in liquid nitrogen and lyophilized to remove the solvent. The lyophilized sample was dissolved in D_2O (99,96% enriched) and left at room temperature overnight in order to replace the exchangeable protein protons with deuterium. The sample was frozen again in liquid nitrogen and D_2O removed by lyophilisation. The procedure was repeated four times in order to ensure the complete exchange. The final solution of the protein (0.6mM) in D_2O was adjusted to
pD 7. 0.5mL of the solution were used in a 10mm tube to measure protein proton relaxation.

For each field of measurement the collective protein proton relaxation rate was calculated by fitting the magnetization decay/recovery curve with the “universal” relaxation rate distribution obtained for well folded proteins [41] reported in Material and Methods. The

fits were quite satisfactory (Figure 2) and reduced χ^2 was sizably smaller (of a factor from 4 to 1.5) than that obtained with monoexponential fits.

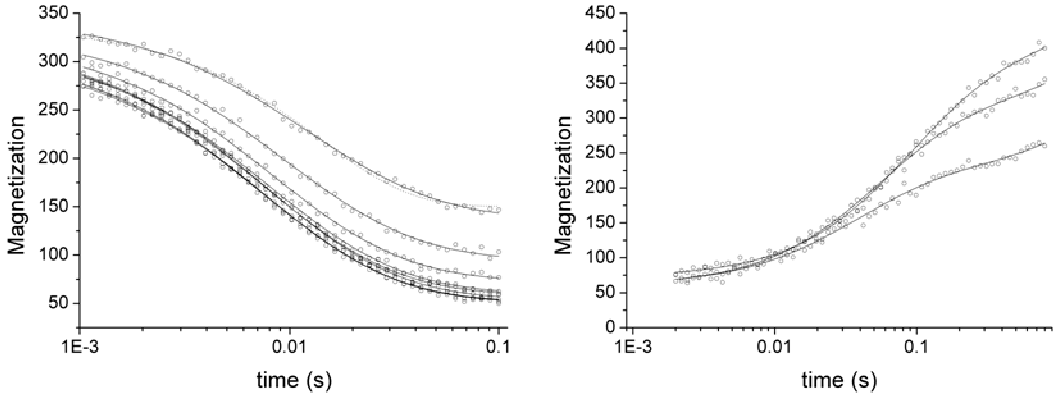


Fig. 2: (Left) Magnetization decays at low magnetic fields (2.8, 1.5, 0.8, 0.4, 0.2, 0.1, 0.07, and 0.04MHz, from top to bottom). The monoexponential fit is shown as a dotted line for the magnetization decay at 2.8 MHz. (Right) Magnetization recovers at high magnetic fields (30, 20, and 14 MHz, from top to bottom). In both panels, best-fit lines were calculated using the ‘‘universal’’ distribution of the relaxation rates defined in Luchinat and Parigi (31) and described in the Materials and Methods section.

The collective relaxation rate values obtained this way and corresponding to the weighted average of the relaxation rates of all non-exchangeable protons in the protein, are reported as a function of field in Figure 3.

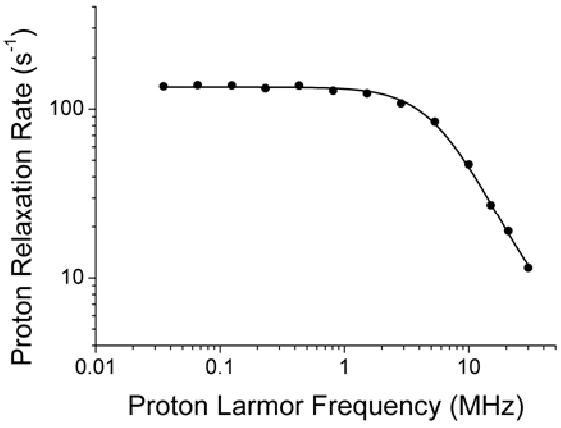


Fig. 3: Collective protein proton relaxation rates for 0.6 mM apocalmodulin, calculated as the weighted average of the relaxation rates obtained from the ‘‘universal’’ distribution (31). The solid line shows the best-fit profile according to Eq. 1.

When the concentration of the protein was decreased to 0.3 or increased to 1.2 mM, no difference in the relaxation profile was observed. This indicates that there is no protein aggregation, because the latter would be reflected in increasing contributions from molecules with a larger reorientation time. The relaxation profile shows one dispersion around 10 MHz; in general, such dispersion is related to a correlation time in the spectral density function which corresponds to the reorientation time of the protein [39,41]. The profile was fit to Eq. 1

$$R_1 = S_C^2 \left\langle E^2 \right\rangle \left(\frac{0.8\tau_R}{1+4\omega^2\tau_R^2} + \frac{0.2\tau_R}{1+\omega^2\tau_R^2} \right) + \alpha \qquad (1)$$

where τ_R, S_C^2 and α are fit parameters. The validity of such equation for the collective protein proton relaxation rates as a function of the field was demonstrated by simulations performed using relaxation data generated with CORMA [41]. However, the same equation (as well as the equation commonly used for unlike spins, due to the onset of the cross relaxation/spin diffusion effects at high fields) would be incorrect for the individual proton relaxation dispersions [41].

From the average of all proton-proton dipolar interactions using the program CORMA [43,44] and the structure of apoCaM (PDB 1QX5), $\langle E^2 \rangle$ was calculated to be 26.0×10^9 s^{-2} (25.4×10^9 s^{-2} if the 1LKJ_1 structure is used). $\langle E^2 \rangle$ can also be obtained from the measurement of the second moment as achievable from the FID of the NMR signal of the protein in the solid state. For largely hydrated proteins the second moment nicely matches with the $S_C^2\langle E^2 \rangle$ value obtained from low field relaxation measurements [40,41,45]. (The second moment of a dry protein can be different from that of the hydrated system due to protein structural changes occurring in the presence of water (30). It has also been showed that hydration decreases the measured second moment because the latter is affected by the motional freedom experienced by the side chains, so that when the high-hydration plateau is reached, it becomes a measure of $S_C^2\langle E^2 \rangle$ rather than of $\langle E^2 \rangle$). The $\langle E^2 \rangle$ value, although dependent on the local accuracy of the protein structural model, because it is a function of the sixth power of the distance between close protons, has been shown to be relatively constant and equal to $(27\pm3)\times10^9$ s^{-2} in a large variety of regularly folded proteins [41]. This is due to the fact that the major contribution to $\langle E^2 \rangle$ is given by the dipolar interactions between protons belonging to the same methylene and methyl groups, which have a fixed distance. The accuracy of the estimate of $\langle E^2 \rangle$ reflects on the accuracy of the determined S_C^2 value.

The best fit value of τ_R is 13.6±0.5 ns, and it represents the correlation time modulating the dipolar interactions, i.e., the reorientation time of the protein (the error is estimated from the standard deviation). This value is in good agreement with the harmonic mean correlation time calculated by HYDRONMR [46] from the anisotropic rotational diffusion tensor using the X-ray structure of apoCaM (PDB 1QX5) in D_2O at 298 K, equal to 14.0 ns. The harmonic mean reorientation time calculated by HYDRONMR using the apoCaM extended solution structure (1LKJ_1) is 19 ns, and it decreases down to 15 ns for other structures reported within the same family where the two domains are differently oriented and closer to one another; that of individual N-terminal (residues 1-80) and C-terminal (residues 81-146) domains is 7.9 and 5.5 ns respectively. Therefore, the experimentally obtained τ_R value indicates that also in solution the reorientation time of apoCaM is basically that of the protein with the two domains in a closed position; this indicates that the protein would preferentially assume one or more compact structures. This is in line with other biochemical and biophysical measurements [47] (see below), and not in contrast with the NMR solution studies. Indeed, the latter indicate the lack of structural interactions between the two protein domains and mobility of the interdomain linker, both phenomena being still consistent with a fluxional behavior of the domains resulting in an ensemble of "closed" conformations.

The best fit value for the squared collective order parameter S_C^2 related to the global protein reorientation time is 0.37±0.02. The accuracy of such parameter (0.02 is the standard deviation from the fit) can be estimated to be about 0.07 as a consequence of the accuracy of the calculated $\langle E^2 \rangle$ value, in the hypothesis of a single correlation time responsible for the observed dispersion. This value is, beyond any experimental uncertainly, much smaller than that found for lysozyme and albumin [39], where S_C^2 was calculated to be about 0.75. This indicates that internal dynamics in apoCaM must be much more effective than in the latter proteins.

The parameter α represents the protein proton collective relaxation value achieved when the τ_R dispersion is completed. Therefore, it is given by the contribution to relaxation of the $(1- S_C^2)$ term times the spectral density function related to local motions and occurring at time scales faster than those observable in the detected field range. From the value of α as obtained from the fit of the experimental profile to Eq. 1 ($\alpha = 4.2 \pm 0.6\ s^{-1}$), correlation time values for the fast local motions of about 0.2 ns are obtained, using the relationship $\alpha = (1- S_C^2) \langle E^2 \rangle \langle \tau_{fast} \rangle$. We note that in this case the value of α is about the double of the value obtained for lysozyme and albumin [39,41], consistent with the much larger $(1- S_C^2)$ value.

^{15}N relaxation studies indicated that the two CaM domains experience a restricted mobility superimposed to the overall tumbling of the molecule [20]. Further analyses were thus performed to investigate whether the reduction in S_C^2 with respect to the values obtained for the other proteins can be ascribed to the motion of the single domains (global internal motion model). Therefore, the relaxation profile of apoCaM was fit using two collective order parameters and two correlation times, with one of them fixed to 6.5 ns, i.e., around that of the individual domains of the protein, to possibly detect the effect of an independent restricted reorietational motion of the two domains, according to Eq. 2.

$$R_1 = \left\langle E^2 \right\rangle \left[S_{C(1)}^2 \left(\frac{0.8\tau_{R1}}{1+4\omega^2\tau_{R1}^2} + \frac{0.2\tau_{R1}}{1+\omega^2\tau_{R1}^2} \right) + S_{C(2)}^2 \left(\frac{0.8\tau_{R2}}{1+4\omega^2\tau_{R2}^2} + \frac{0.2\tau_{R2}}{1+\omega^2\tau_{R2}^2} \right) \right] + \alpha \tag{2}$$

The fit was equally good and a first reorientation time of 14.4 ns is calculated with $S_{C(1)}^2$ equal to 0.32; the $S_{C(2)}^2$ corresponding to the motion at 6.5 ns was calculated to be 0.06, i.e. much smaller (see Table 1).

	Isotropic model		Global internal motion model			
	τ_R (ns)	S_C^2	τ_{R1} (ns)	$S_{C(1)}^2$	τ_{R2} (ns)	$S_{C(2)}^2$
Low-field relaxation (this work)	13.6	0.37	14.4	0.32	6.5 (fixed)	0.06
			14.0	0.35	3.7 (fixed)	0.05
High-field relaxation(14)	9.1–10.3*	0.85	11.8–16.7*	0.53–0.60	2.7–4.2*	0.25–0.32

*Values are scaled by the factor 1.11/0.9 = 1.23 for comparison with the values in D_2O used in this study.

Table 1: Summary of best-fit correlation times and order parameters obtained for apoCaM.

The presence of multiple correlation times cannot thus increase the S_C^2 value corresponding to the slowest motion. This proves that the reduction in the S_C^2 value observed for CaM with respect to the other proteins is indeed almost independent on the number of correlation times used to fit the data, and that the value itself is quite robust. It also proves that the fastest motions predominantly occur on time scales shorter than the time expected for the reorientation of the single domains.

5.4 Discussion

The reorientation time, calculated with this work, reveals that for most of the time the protein posses a relatively compact structure, and thus that the two domains must be close to one another even in solution, although the absence of interdomain NOEs [2] suggests that a dynamic ensemble of conformations must be experienced. This result is relevant and is in line with the previously found interdomain contacts in apoCaM in solution, through fluorescence spectroscopy measurements performed on engineered apoCaM mutants [48]. Moreover, fluorescence resonance energy transfer (FRET) measurements [49] suggest a reorientation time that is in agreement with the overall protein rotational motion, thus indicating that the inter-domain conformational heterogeneity detected through high field NMR measurements must be the result of a number of slowly interconverting (in the nanosecond time scale) distinct conformations [49].

^{15}N relaxation studies previously performed [20] indicated a shorter reorientation time when an isotropic model for protein tumbling was assumed, shorter than expected for a fully compact closed structure. This made the global internal motion model, where internal domains have a restricted mobility superimposed to the overall tumbling of the molecule, preferable, in addition to the significantly better agreement between experimental and best fit values achieved with such model. The global reorientation times obtained from the high field relaxation measurements are largely affected by the selected groups of residues included in the fit and by the fields of measurement, and their uncertainty is thus relatively large. In the case of apoCaM, they span values from 9.6 to 13.6 ns in H_2O [20]. Scaled to D_2O, these values span from 11.8 to 16.7 ns, which compare well with the present best-fit τ_R value of 14 ns in this study. (Table 1).

The fit of the high field relaxation data also indicates the occurrence of internal motions with correlation time of about 3 ns [20]. The latter value seems to be too short for the tumbling of the CaM single domains, which are expected to be around 5-5.5 ns in H_2O. The fit of the present data actually indicates that, if a motion with such correlation time is present, its order parameter must be quite low (see table 1). On the other hand, a fit of equally good quality was also obtained by fixing the internal correlation time to 3.7 ns (corresponding to 3 ns in H_2O) as found in the ^{15}N relaxation study, the other parameters assuming values comprised between those calculated in the isotropic model and in the internal motion model with τ_{R2} fixed to 6.5 ns (see table 1). In conclusion, the present data are not sensitive enough

to determine the presence of motions with correlation times of the order of some nanoseconds, and they are consistent with the presence of such motions only with collective squared order parameters smaller than 0.1. For the purpose of the present study, it is important to note that the presence of such internal motions does not sizably change the value of the overall tumbling and of the corresponding order parameter.

As expected, the obtained S_C^2 values are smaller than the available S^2 values obtained from relaxation rate measurements on backbone nuclei at high fields [26] both in the assumption of an isotropic model with a single reorientation time (S^2=0.85) or of a global internal motion model, described by two order parameters, one for the global reorientation and one for internal protein motions occurring on an intermediate time scale between the global reorientation time and the faster internal dynamics. In the latter case, in fact, an averaged $S_{NH(1)}^2$ value of 0.56 was calculated from the amide relaxation with a correlation time corresponding to the overall reorientation time of the protein [20], and further $S_{NH(2)}^2$ values of 0.25-0.30 for a correlation time of 2.5-3 ns (see table 1). These values are both significantly larger (outside the error) than the $S_{C(1,2)}^2$ values of 0.32 and 0.06 obtained from the fit of the present data.

The comparison of the order parameters determined from high field data and the collective S_C^2 values determined here actually shows that the former does not monitor all motional features present in the protein in the picosecond to nanosecond range. This is because side chains dynamics is not revealed. On the other hand, side chain dynamics may be fundamental in the characterization of the protein. For instance, the low sensitivity of ^{15}N relaxation measurements to side-chain and protein domain motions is for instance indicated by the fact that the S_{NH}^2 values obtained for the calcium-bound CaM in the absence and presence of bound peptides are essentially the same for all residues except those in the linker between the N-terminal and C-terminal domains, despite the fact that the difference in the dynamics within the protein is outstanding [6,23]. Analogously, only slightly different squared order parameters were calculated for the ^{13}CO-$^{13}C\alpha$ vectors for free and complexed CaM [6].

Side-chain mobility is better addressed by the model-free generalized order parameter, S_{axis}^2, for the symmetry axis of methyl groups [50]. From the methyl order parameters observed in calcium-bound CaM, three distinct classes of motions were suggested, centered at squared order parameters of 0.35, 0.58 and 0.78, and peptide binding was shown to increase substantially the S_{axis}^2 measured for some methyl groups, by often moving methyls from one lower order parameter group to a larger order parameter group [24,50]. These works clearly show

that CaM has a peculiar enhanced mobility in the side chains. As suggested in [25], this may be due to the hydrophobic residues allowing much larger side chain mobility than in the side chains of canonical globular proteins. In our study, the collective S_C^2 value found for apoCaM (0.37), dramatically smaller than the value observed for other well folded globular proteins (0.75) [39,41], is even on the lowest side of the range of the S_{axis}^2 values for methyl groups [50]. In this context it is particularly significant that the S_C^2 values provided by the present approach arise from *all* protein protons, 3/4 of which belong to side chains, rather than from methyl protons, which constitute about 1/4 of the total. The strikingly small S_C^2 value obtained in the present work points to an extraordinary mobility not only of methyl groups but of all side chain protons in general.

Such high mobility of side chain protons is not shared by backbone protons, as it results from ^{15}N relaxation data. This makes possible that the "universal" distribution of relaxation rates that was used to fit the data, and which indeed provided a good fit of the magnetization decay/recovery curves, may not approximate the real distribution as satisfactorily as it does for more rigid globular proteins. A fit of the magnetization decay/recovery curves was thus also performed using a double exponential function instead of the "universal" rate distribution, and also in this case the agreement with the experimental data was very good, the resulting relaxation rates being reduced of about 10%. Therefore, the fit of the resulting profile provides basically the same best-fit parameters, the S_C^2 value being only slightly smaller.

Relaxation measurements performed on apoCaM showed that the chemical exchange contributions are substantial for most residues of the C-terminal domain, thus implying intra-domain exchange between conformational substates [26]. This intra-domain conformational exchange appears to involve transitions between a predominantly populated closed conformation (with an antiparallel helical arrangement) and a smaller population of more open conformations (not parallel helical arrangements) of the C-terminal domain [2,10,26,51-55]. A dynamic equilibrium involving conformations with a partially exposed hydrophobic core provides CaM with the ability to interact with its targets in the absence of excess calcium(II) [56,57]. Therefore, EF-hand motifs in apoCaM exist in both closed and open states, possibly sampling a very large spectrum of conformational states, with a conformational exchange rate on microsecond to millisecond time scale and predominance of the closed conformation. Although the generalized order parameter S_C^2 here obtained only reflects motions in a range of time scales from the picoseconds to the nanoseconds, the presence of a large mobility in the

latter scale is likely to be associated with mobility in a slower scale as well.

Large mobility seems to be present also in the N-terminal domain, and not only within the residues corresponding to the calcium binding loops. It was in fact shown that the first helix of the N-terminal domain of apoCaM undergoes large amplitude nanosecond motions [58]. This indicates a considerable dynamic flexibility of the first helix of the N-terminal domain, and thus within the domain, in apoCaM. This is in line with molecular dynamics simulations that predicted the presence of four subdomains within the amino-terminal domain of the protein that have the potential to undergo large amplitude independent motions relative to one another [4].

In conclusion, the measurements performed on apoCaM in this work provide direct information on both the reorientation time of the investigated system and on the extent of internal motions of protein protons. We have found that, i) the reorientation time obtained from the fit of the spectral density function of protein protons indicates that the protein is mainly in a conformation with the two domains in close contact, and ii) the collective squared order parameter obtained in the present analysis is much smaller than that obtained for other globular proteins. Such low S_C^2 value depends on the larger side chain dynamics of apoCaM, because 3/4 of the protons contributing to the S_C^2 value belong to side chains, and backbone nuclei are known to have a regular mobility from ^{15}N high field relaxation measurements. Therefore, S_C^2 beautifully complements other measurements to describe the extent of both global and local the motional features experienced by the system.

Protein proton relaxation measurements at low fields, providing direct access to the spectral density function thus allow us to safely recover the reorientation time of the protein, which at high fields may result less straightforward to obtain because the corresponding dispersion is already largely occurred. From the reorientation time value, further information can be derived on the conformational state of the investigated systems. They also provide a collective order parameter which monitors side chain mobility, not accounted in standard high field ^{15}N relaxation measurements. These findings point out some global aspects of the motion that complement other aspects available from different techniques, so that all together a complete picture for the protein dynamics can be recovered. The approach here proposed can be even more relevant for mobility studies of large proteins, when information from high field ^{15}N T_1, T_2, NOE measurements can hardly be obtained due to severe line broadening of the NMR spectra, as well as for the investigation of dynamics in protein-protein adducts.

Reference List

1. Barbato,G., Ikura,M., Kay,L.E., Pastor,R.W. & Bax,A. Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy; the central helix is flexible. Biochemistry 31, 5269-5278 (1992).

2. Kuboniwa,H. et al. Solution structure of calcium-free calmodulin. Nature Struct. Biol. 2, 768-776 (1995).

3. Evenäs,J., Forsén,S., Malmendal,A. & Akke,M. Backbone dynamics and energetics of a calmodulin domain mutant exchanging between closed and open conformations. J. Mol. Biol. 289, 603-617 (1999).

4. Wriggers,W., Mehler,E., Pitici,F., Weinstein,H. & Schulten,K. Structure and Dynamics of Calmodulin in Solution. Biophys. J. 74, 1622-1639 (1998).

5. Bertini,I. et al. Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. Proc. Natl. Acad. Sci. USA 101, 6841-6846 (2004).

6. Wang,T., Frederick,K.K., Igumenova,T.I., Wand,A.J. & Zuiderweg,E.R.P. Changes in calmodulin main-chain dynamics upon ligand binding revealed by cross-correlated NMR relaxation measurements. J. Am. Chem. Soc. 127, 828-829 (2005).

7. Bertini,I. et al. Paramagnetism-Based NMR Restraints Provide Maximum Allowed Probabilities for the Different Conformations of Partially Independent Protein Domains. J. Am. Chem. Soc. 129, 12786-12794 (2007).

8. Schumacher,M.A., Crum,M. & Miller,M.C. Crystal Structures of Apocalmodulin and Apocalmodulin/SK Potassium Channel Gating Domain Complex. Structure 12, 849-860 (2004).

9. Babu,Y.S., Bugg,C.E. & Cook,W.J. Structure of calmodulin refined at 2.2 Å resolution. J. Mol. Biol. 204, 191-204 (1988).

10. Zhang,M., Tanaka,T. & Ikura,M. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. Nat Struct Biol 2, 758-767 (1995).

11. Lange,O.F. et al. Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. Science 320, 1471-1475 (2008).

12. Valentine,E.R. & Palmer,A.G. Microsecond-to-millisecond conformational dynamics demarcate the GluR2 glutamate receptor bound to agonists glutamate, quisqualate, and AMPA. Biochemistry 44, 3410-3417 (2005).

13. Ryabov,Y.E. & Fushman,D. A model of Interdomain Mobility in a Multidomain Protein. J. Am. Chem. Soc. 129, 3315-3327 (2007).

14. Ishima,R. & Torchia,D.A. Protein dynamics from NMR. Nature Struct. Biol. 7, 740-743 (2000).

15. Mulder,F.A.A., Mittermaier,A., Hon,B., Dahlquist,F.W. & Kay,L.E. Studying excited states of proteins by NMR spectroscopy. Nature Struct. Biol. 8, 932-935 (2001).

16. Kay,L.E. NMR studies of protein structure and dynamics. J. Magn. Reson. 173, 193-207 (2005).

17. Fischer,M.W.F., Zeng,L., Majumdar,A. & Zuiderweg,E.R.P. Characterizing semilocal motions in proteins by NMR relaxation studies. Proc. Natl. Acad. Sci. USA 95, 8016-8019 (1998).

18. Brüschweiler,R. New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins. Curr. Opin. Struct. Biol. 13, 175-183 (2003).

19. Baber,J.L., Szabo,A. & Tjandra,N. Analysis of Slow Interdomain Motion of Macromolecules Using NMR Relaxation Data. J. Am. Chem. Soc. 123, 3953-3959 (2001).

20. Tjandra,N., Kuboniwa,H., Ren,H. & Bax,A. Rotational dynamics of calcium-free calmodulin studied by 15N-NMR relaxation measurements. Eur. J. Biochem. 230, 1014-1024 (1995).

21. Akke,M. & Palmer,A.G., III. Monitoring macromolecular motions on microsecond to millisecond time scales by R1rho-R1 constant relaxation time NMR spectroscopy. J. Am. Chem. Soc. 118, 911-912 (1996).

22. Fushman,D., Varadan,R., Assfalg,M. & Walker,O. Determining domain orientation in macromolecules by using spin-relaxation and residual dipolar coupling measurements. Progr. NMR Spectrosc. 44, 189-214 (2004).

23. Igumenova,T.I., Lee,A.L. & Wand,A.J. Backbone and Side Chain Dynamics of Mutant Calmodulin-Peptide Complexes. Biochemistry 44, 12627-12639 (2005).

24. Frederick,K.K., Marlow,M.S., Valentine,K.G. & Wand,A.J. Conformational entropy in molecular recognition by proteins. Nature 448, 325-329 (2007).

25. Lee,A.L., Kinnear,S.A. & Wand,A.J. Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. Nat. Struct. Biol. 7, 72-77 (2000).

26. Malmendal,A., Evenäs,J., Forsén,S. & Akke,M. Structural dynamics in the C-terminal domain of calmodulin at low calcium levels. J. Mol. Biol. 293, 883-899 (1999).

27. Urbauer,J.L., Short,J.H., Dow,L.K. & Wand,A.J. Structural Analysis of a Novel Interaction by Calmodulin: High-Affinity Binding of a Peptide in the Absence of Calcium. Biochemistry 34, 8099-8109 (1995).

28. Ferrante,G. & Sykora,S. Technical aspects of fast field cycling. Adv. Inorg. Chem. 57, 405-470 (2005).

29. Kimmich,R. & Anoardo,E. Field-cycling NMR relaxometry. Progr. NMR Spectrosc. 44, 257-320 (2004).

30. Aime,S. et al. Relaxometric evaluation of novel manganese(II) complexes for

application as contrast agents in magnetic resonance imaging. J. Biol. Inorg. Chem. 7, 58-67 (2002).

31. Kowalewski,J. & Maler,L. Nuclear Spin Relaxation in Liquids: Theory, Experiments, and Applications. Taylor & Francis, (2006).

32. Kowalewski,J., Kruk,D. & Parigi,G. NMR relaxation in solution of paramagnetic complexes: recent theoretical progress for S>1. Adv. Inorg. Chem. 57, 41-104 (2005).

33. Schühle,D.T. et al. Calix[4]arenes as Molecular Platforms for Magnetic Resonance Imaging (MRI) Constrast Agents. Chemistry 15, 3290-3296 (2009).

34. Assfalg,M. et al. NMR Structural Studies of the Supramolecular Adducts between a Liver Cytosolic Bile Acid Binding Protein and Gadolinium(III)-Chelates Bearing Bile Acids Residues: Molecular Determinants of the Binding of a Hepatospecific Magnetic Resonance Imaging Contrast Agent. J. Med. Chem. 50 , 5257-5268 (2007).

35. Rinck,P.A., Fischer,H.W., Vander Elst,L., Van Haverbeke,Y. & Muller,R.N. Field-cycling relaxometry: medical applications. Radiology 168, 843-849 (1988).

36. Major,J.L., Meade,T.J., Parigi,G. & Luchinat,C. The Synthesis and In-Vitro testing of a Zn(II)-Activated MR contrast agent. Proc. Natl. Acad. Sci. USA 104, 13881-13886 (2007).

37. Urbanczyk-Pearson,L.M. et al. Mechanistic Investigation of b-galactosidase-activated MR Contrast Agents. Inorg. Chem. 47, 56-68 (2008).

38. Caravan,P. et al. Albumin binding, relaxivity and water exchange kinetics of the diastereoisomers of MS-325 a gadolinium(III) based magnetic resonance angiography contrast agent. Inorg. Chem. 46, 6632-6639 (2007).

39. Bertini,I. et al. NMR spectroscopic detection of protein protons and longitudinal relaxation rates between 0.01 and 50 MHz. Angew. Chem. Int. Ed. 44, 2223-2225 (2005).

40. Diakova,G., Goddard,Y.A., Korb,J.-P. & Bryant,R.G. Changes in protein structure and dynamics as a function of hydration from 1H second moments. J. Magn. Reson. 189, 166-172 (2007).

41. Luchinat,C. & Parigi,G. Collective relaxation of protein protons at very low magnetic field: a new window on protein dynamics and aggregation. J. Am. Chem. Soc. 129 , 1055-1064 (2007).

42. Korb,J.-P. & Bryant,R.G. The physical basis for the magnetic field dependence of proton spin-lattice relaxation rates in proteins. J. Chem. Phys. 115, 10964-10974 (2001).

43. Keepers,J.W. & James,T.L. A theoretical study of distance determinations from NMR. Two-dimensional nuclear Overhauser effect spectra. J. Magn. Reson. 57, 404-426 (1984).

44. Borgias,B., Gochin,M., Kerwood,D.J. & James,T.L. Relaxation matrix analysis of 2D NMR data. Progr. NMR Spectrosc. 22, 83-100 (1990).

45. Bertini,I., Luchinat,C. & Parigi,G. Solution NMR of paramagnetic molecules. Elsevier, Amsterdam (2001).

46. Garcìa de la Torre,J., Huertas,M.L. & Carrasco,B. HYDRONMR: Prediction of NMR Relaxation of Globular Proteins from Atomic-Level Structures and Hydrodynamic Calculations. J. Magn. Reson. 147, 138-146 (2000).

47. Qin,Z. & Squier,T.C. Calcium-Dependent Stabilization of the Central Sequence Between Met76 and Ser81 in Vertebrate Calmodulin. Biophys. J. 81, 2908-2918 (2001).

48. Boschek,C.B., Squier,T.C. & Bigelow,D.J. Disruption of Interdomain Interactions via Partial Calcium Occupancy of Calmodulin. Biochemistry 46, 4580-4588 (2007).

49. Sun,H., Yin,D. & Squier,T.C. Calcium-Dependent Structural Coupling between Opposing Globular Domains of Calmodulin Involves the Central Helix. Biochemistry 38, 12266-12279 (1999).

50. Marlow,M.S. & Wand,A.J. Conformational dynamics of calmodulin in complex with the calmodulin-dependent kinase kinase calmodulin-binding domain. Biochemistry 45, 8732-8741 (2006).

51. Evenas,J., Malmendal,A., Thulin,E., Carlström,G. & Forsén,S. Ca2+ binding and conformational changes in a calmodulin domain. Biochemistry 37, 13744-13754 (1998).

52. Capozzi,F., Luchinat,C., Micheletti,C. & Pontiggia,F. Essential dynamics of helices provide a functional classification of EF-hand proteins. J. Proteome Res. 6, 4245-4255 (2007).

53. Babini,E. et al. Principal component analysis of a comprehensive structural database of EF-hand domains to describe the conformational freedom within the EF-hand superfamily. J. Proteome Res. 4, 1961-1971 (2005).

54. Houdusse,A., Silver,M. & Cohen,C. A model of Ca2+-free calmodulin binding to unconventional myosins reveals how calmodulin acts as a regulatory switch. Structure 4, 1475-1490 (1996).

55. Evenäs,J., Malmendal,A. & Akke,M. Dynamics of the Transition between Open and Closed Conformations in a Calmodulin C-Terminal Domain Mutant. Structure 9, 185-195 (2001).

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

57. Houdusse,A. et al. Crystal structure of apo-calmodulin bound to the first two IQ motifs of myosin V reveals essential recognition features. Proc. Natl. Acad. Sci. USA 103, 19326-19331 (2006).

58. Chen,B., Mayer,M.U., Markillie,L.M., Stenoien,D.L. & Squier,T.C. Dynamic Motion of Helix A in the Amino-Terminal Domain of Calmodulin Is Stabilized upon Calcium Activation. Biochemistry 44, 905-914 (2005).

Chapter 6

Entropic contribution to the linking coefficient in Fragment Based Drug Design: a case study

Borsi, V., Calderone, V. , Fragai, M., Luchinat, C. and Sarti N.

Submitted

6.1 Introduction

Several high-throughput methods have been developed over the years in order to speed-up the screening and the identification of new hits for pharmaceutical targets[1-10,10-12]. Usually, screening leads to identify molecules that need subsequent optimization steps in order to reach a satisfactory affinity. Among the screening methods, the fragment-based approach has seen an unquestionable success in both pharmaceutical industry and academic setting due to its high hit rate in generating candidate drugs[1,12-17].

The leading concept of the fragment-based approach is that a high affinity compound can be designed by tethering, with a suitable linker, two or more (even weak) ligands able to bind adjacent protein sites[1]. As it is easier to optimize the binding of a small fragment than of a large molecule, linking to separately optimized fragments has a better chance to produce compounds with nanomolar dissociation constant K_D. As an added value, in several cases the K_D of the tethered molecules is found to be smaller than the product of the K_Ds of the single fragments[18,19]. This additional increase in affinity can be explained by considering that, when two ligands bind to a protein, each of them loses a fraction of rigid body rotational and translational entropy. Conversely, when the two ligands are tethered in a single molecule, only one unfavourable rigid body entropy barrier affects the binding[20]. This smaller loss of entropy (i.e. relative entropic gain) accounts for the linking effect[21,22]. For a two-fragment A-B molecule, the dissociation constant K_D^{AB} is often reported as the product of the dissociation constants of the isolated fragments A and B, K_D^{A} and K_D^{B}, multiplied by a term named *linking coefficient*, E[1,23]:

$$K_D^{AB} = K_D^{A} \cdot K_D^{B} \cdot E \quad \text{(or, equivalently,} \qquad \Delta G^{AB} = \Delta G^{A} + \Delta G^{B} + RT \ln E) \quad (1)$$

where $E < 1$ implies a favourable contribution to binding.

In an ideal case, the linking coefficient E, obtained from (1) when the three binding constants are know from experiments, would thus be a direct measure of the entropic gain.

However, besides the favorable entropic contribution, there may be unfavorable enthalpic contributions to the binding that arise from distortions of the target binding geometry of the two fragments when linked together, as well as either favourable or unfavorable interactions of the linker itself with the target molecule. Indeed, it has been

observed that the linking coefficient is strongly dependent on the structural features of the tethered molecule and on its binding mode to the target. In some cases linking coefficients even larger than one (i.e. unfavorable) have been found, while in successful cases linking coefficient values slightly smaller than 10^{-2} have been observed[19]. The most favourable *E* value (3.1×10^{-3} M^{-1}) has been observed for a biphenyl-based inhibitor of matrix metalloproteinase 2 (MMP-2) [19]. Despite the pure entropic term would be theoretically capable to account by itself for linking coefficients as low as 10^{-8} M^{-1}, [20] experimental verification that even *E* values of the order of 10^{-2} M^{-1} are entirely due to a favorable entropic term is still lacking.

One way to obtain this verification would be to construct a two-fragment molecule where the two fragments are able to bind separately to the target in such a way that they just need to be linked together by one covalent bond without any intervening atom. The increase in binding affinity of a zero-length tether molecule of this type should be, at least to a first approximation, entirely due to the entropic effect, provided that the enthalpic contributions are essentially unchanged, i.e. the two fragments maintain all relevant interactions with the target when bound together. Interestingly, at least one case of this type has been described, and a reasonable although not extremely small *E* value of about 3.3×10^{-2} M^{-1} has been reported[24]. Unfortunately, experimental verification of its entropic nature through e.g. calorimetry is not available.

Here we report on a sulfonamide derivative inhibitor of matrix metalloproteinases (MMP) whose components have been designed by application of a *deconstructing* procedure[25], and where the linker is just one covalent bond. Sulfonamide derivatives constitute one of the most important classes of inhibitors of zinc enzymes, possessing a high affinity towards several zinc enzyme families including, besides MMPs, also carbonic anhydrases[26-30]. In the case of MMPs, sulphonamide inhibitors are constituted by a zinc binding group such as hydroxamate or carboxylate and by a lipophilic moiety interacting with the S1' cavity[31]. N-hydroxy-2-(4-methoxyphenylsulfonamido) acetamide (PMAHA, Figure 1) is a simple but relatively potent inhibitor of MMP-12[32,33]. It can be easily synthesized by reacting the glycine methyl-ester with paramethoxy-sulphonyl chloride and then by deprotecting the methyl-ester to form the hydroxamic moiety. Virtually, this molecule can be considered as formed by two fragments, paramethoxybenzene-sulfonyl amide (PMS) and acetohydroxamic acid (AHA) (Figure 1), tethered by a simple covalent bond upon removal of one hydrogen atom on each side. The present case study analyzes the binding of PMAHA, as

well as of its constituting fragments PMS and AHA, to the catalytic domain of MMP-12 (cdMMP-12). The system displays a very small value of E of 2.1×10^{-3} M^{-1}, and it is shown that this value is entirely due to a very favorable entropic contribution, the enthalpic term being very similar to the sum of those of the two isolated fragments.

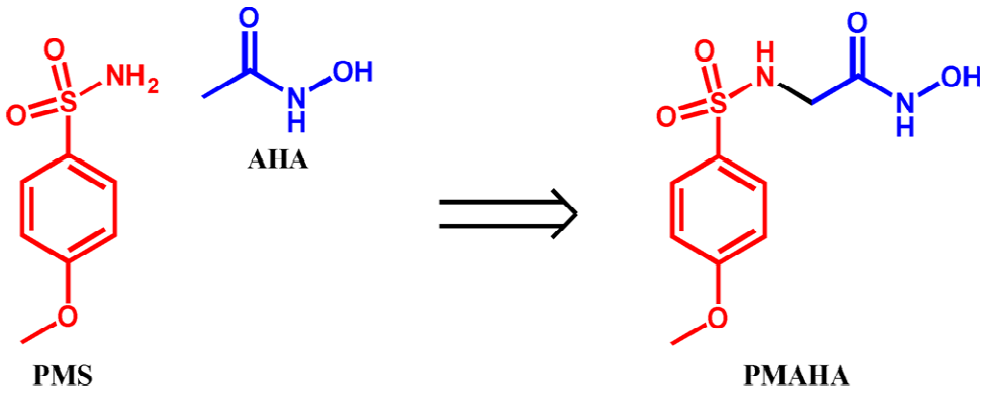


Fig. 1: N-hydroxy-2-(4-methoxyphenylsulfonamido)acetamide (PMAHA) can be considered as formed by two fragments, paramethoxybenzene-sulfonyl amide (PMS) and acetohydroxamic acid (AHA).

6.2 Material and methods

Sample preparation. The cDNA encoding the cdMMP-12 (Gly106–Gly263) was amplified by a polymerase chain reaction (PCR) from IMAGE consortium clone (ID 196612) using two synthetic oligonucleotides as primers. The cDNA obtained was cloned into the pET21a vector (Novagen) between the restriction sites NdeI and BamHI. The single amino acid substitution, to obtain the F171D mutant, was created using the QuickChange™ Site-Directed Mutagenesis Kit from Stratagene. The construct was transformed into the BL21Gold (DE3) strain for expression of recombinant protein. The cells were grown in 2×YT media at 37 °C. and the protein expression was induced during the exponential growth phase with 0.5 mM IPTG. Cells were left grow for 4 h after induction and were harvested by centrifugation. After lysis of the cells the inclusion bodies, containing the protein, were solubilized in 8 M urea and 20 mM sodium acetate (pH 5.0). The protein was purified on the Hitrap SP column (Pharmacia) with a buffer containing 6 M urea and 20 mM sodium acetate (pH 5.0). The elution was performed using a linear gradient of NaCl up to 0.35 M. The purified protein was refolded by using a multistep dialysis against a solution containing 50 mM Tris–HCl (pH 7.2),

10 mM $CaCl_2$, 0.1 mM $ZnCl_2$, 300 mM NaCl, 500 mM acetohydroxamic acid (AHA), and decreasing concentration of urea (from 4 to 2 M). The last two dialyses were performed against a solution containing 20 mM Tris–HCl (pH 7.2), 5 mM $CaCl_2$, 0.1 mM $ZnCl_2$, 300 mM NaCl, and 200 mM AHA. The final yield of the purified protein was approximately 30mg/l.

Protein expression was also carried out in minimal medium containing 1.25 g/l of $(^{15}NH_4)_2SO_4$ in order to produce ^{15}N-enriched protein. Typically, about 22 mg of enriched protein could be obtained from a 1 l culture.

NMR spectroscopy. $^{1}H–^{15}N$ HSQC experiments, implemented with the sensitivity enhancement scheme, were recorded at 298 K on a Bruker Avance 700 operating at proton nominal frequency of 700.21 MHz. To determine the binding affinity of PMS towards the catalytic domain of MMP-12, the alteration of the chemical shifts induced on 2D ^{1}H-^{15}N HSQC spectra upon the titration with PMS was monitored by NMR. The experiments were performed on samples containing 0.289 mM of ^{15}N-enriched MMP-12 in presence of 10 mM Tris–HCl buffer (pH 7.2), 10 mM $CaCl_2$, 0.1 mM $ZnCl_2$, 300 mM NaCl with and without 200 mM acetohydroxamic acid. The dissociation constant was calculated by fitting the $\Delta\delta$ as a function of the ligand concentration.

X-Ray measurement. Crystals of human MMP12, already containing AHA from the refolding process, grew at 20 °C from a 0.1 M Tris-HCl, 30% PEG 6000, 200 mM AHA, 1.0 M $LiCl_2$ solution at pH 8.0 using the vapor diffusion technique. The final protein concentration was about 10 mg/ml. A soaking procedure was carried out to obtain the adduct with PMS. The data were measured in-house, using a PX-Ultra copper sealed tube source (Oxford Diffraction) equipped with an Onyx CCD detector. The dataset was collected at 100 K and the crystal used for data collection was cryo-cooled without any cryo-protectant treatment. The crystal diffracted up to 1.8 Å resolution and belongs to spacegroup C2 (a= 51.57 Å b= 60.19 Å c= 54.21 Å, α= γ= 90° β= 115.09°) with one molecule in the asymmetric unit, a solvent content of about 50% and a mosaicity of 0.7°. The data were processed in all cases using the program MOSFLM[34] and scaled using the program SCALA[35] with the TAILS and SECONDARY corrections on (the latter restrained with a TIE SURFACE command) to achieve an empirical absorption correction. Table 2S shows the data collection and processing statistics for all datasets. The structure was solved using the molecular replacement technique; the model used was that of the MMP12-AHA adduct (1Y93) from where the inhibitor, all the

water molecules and ions were omitted. The correct orientation and translation of the molecule within the crystallographic unit cell was determined with standard Patterson search techniques[36,37] as implemented in the program MOLREP[38,39]. The isotropic refinement was carried out using REFMAC5[40] but the metal ions B-factors were refined taking anisotropy into account. In between the refinement cycles the models were subjected to manual rebuilding by using XtalView[41]. The same program has been used to model all inhibitors. Water molecules have been added by using the standard procedures within the ARP/WARP suite[42] and for the atomic resolution datasets hydrogens were added at the riding positions and refined. The stereochemical quality of the refined models was assessed using the program Procheck[43]. The Ramachandran plot is of very good quality. Table S1 reports the data collection and refinement statistics.

Calorimetry. Isothermal Titration Microcalorimetry experiments were performed at 298 K with a VP-ITC microcalorimeter (MicroCal, Inc., Northampton, MA). After an initial injection of 1µL, aliquots of 10µL of 200µM BMAHA inhibitor with 0.1% (v/v) DMSO were stepwise injected into the sample cell containing a solution 20µM of MMP12 catalytic domain until complete saturation. All experiments were performed in 20mM Tris (pH 7.2), 5mM $CaCl_2$, 0.1mM $ZnCl_2$, 0.3M NaCl, AHA 4mM with PMS in concentrations ranging from 0.1 to 3 mM. Heats of dilution were measured by injecting the ligand into buffer and then subtracted from the binding heats. The thermodynamic parameters and K_A values were calculated fitting data to a single binding site model with ORIGIN 7.0 sofware (Microcal, Inc.).

6.3 Results

The interaction of PMS with the active site of MMP-12 was analyzed by X-ray crystallography, NMR and microcalorimetry. The dissociation constant of PMS was obtained from the analysis of the chemical shifts perturbations on the assigned ^{1}H-^{15}N HSQC spectra of the free and AHA-inhibited form of MMP-12. The value of the dissociation constant (K_D = 1.5×10^{-3} M) was not affected by the presence of different concentrations of AHA in solution. The binding mode of the ligand in the active site of the protein was investigated by soaking crystals of MMP-12, incorporating the weak AHA inhibitor, in solutions of PMS.

In the X-ray structure, the PMS ligand fits the S1' cavity with the p-methoxybenzene group and establishes an H-bond between one of its sulfonyl oxygens (O2) and the amide

nitrogen of Leu-181 (2.93 Å) (Figure 2B). At the same time, the free AHA ligand chelates the zinc ion with its two oxygens, and the interaction is further stabilized by two H-bonds, one between the protonated oxygen atom (O4) of AHA and the carboxylate Oδ2 of Glu-219 (2.53 Å) and the second (2.76 Å) between the hydroxamate NH and the carbonyl oxygen of Ala-182. It should be pointed out that the two ligands bind independently to the protein, and do not interact with one another. Moreover, the AHA ligand exhibits the same binding mode in the presence and in the absence[32] of PMS.

The binding of PMAHA (Figure 2C) and AHA alone (Figure 2A) with MMP-12 was characterized previously by X-ray, microcalorimetry and enzymatic assay[32,33]. In the complex of MMP-12 with PMAHA, the inhibitor binds the catalytic zinc ion by its hydroxamic moiety, and its p-methoxybenzene group fits the S1' cavity.

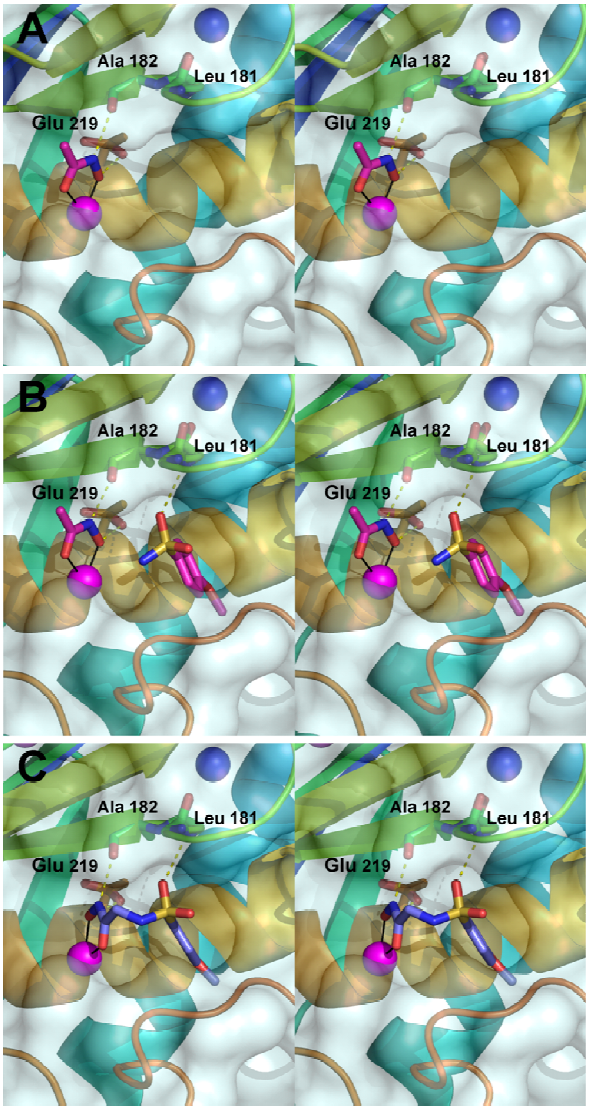


Fig. 2: Stereo view of the active site of MMP-12 complexed with hydroxamic acid (**A**). PMS and AHA bind independently to the active site of MMP-12 protein and do not interact with one another (**B**). The presence of PMS do not alter the binding mode of AHA. The tethered molecule PMAHA maintains all the interactions observed for the two isolated fragments (**C**). Only the AHA moiety appears rotated with respect to the position of the isolated AHA ligand, but with all relevant protein interactions maintained.

The paramethoxy-sulfonyl moiety in the PMAHA molecule shares the same binding mode of the isolated PMS ligand, while the AHA moiety appears rotated by ca. 45° with respect to the position of the isolated AHA ligand[32]. Despite this rotation, all the interactions observed for the two fragments are in place: the bidentate behavior of the AHA moiety, the sulfonyl oxygen at 2.93 Å from the amide nitrogen of Leu-181, the protonated oxygen atom (O4) at 2.53 Å from the carboxylate Oδ2 of Glu-219, and the hydroxamate NH at 2.76 Å from the Ala-182 carbonyl oxygen. A series of ITC measurements on the binding of a parent arysulfonamide inhibitor (BSAHA, see Figure 3) were performed in the presence of various PMS concentrations in the range 0.1 to 3 mM.

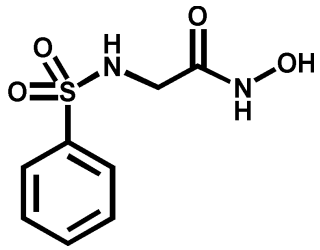


Fig. 3: Structure of N-hydroxy-2-(phenylsulfonamido)acetamide (BSAHA).

AHA was also present in solution at a concentration of 4 mM. By fitting the obtained enthalpy as a function of PMS concentration and after the correction for the competition with AHA, a ΔH^0 of -6.0 kcal/mol was found for the binding of PMS. The ΔG^0, ΔH^0, ΔS^0 and K_D values for the binding of AHA and PMAHA to MMP-12 at 298 K were determined previously[33] and are reported in Table 1. For PMS the free energy of binding (ΔG^0) was calculated from the dissociation constant obtained by NMR. The ΔS^0 value was estimated from the relationship $\Delta G^0 = \Delta H^0 - T\Delta S^0$. The value of ΔH_0 for PMS was determined by isothermal titration microcalorimetry (ITC) in this work. As it appears from Table 1, the value of ΔH^0 found for PMAHA (-8.52 kcal/mol) is similar to the sum of the binding enthalpy of AHA and PMS. At the same time the free energy of binding for PMAHA is much more negative than the sum of the ΔG^0 of the isolated fragments. The resulting linking coefficient is about 2.1×10^{-3} M^{-1} (see Table 1).

	K_D (M)	ΔG^0 (kcal/mol)	ΔH^0 (kcal/mol)	$-T\Delta S^0$ (kcal/mol)
AHA	$6.2 \cdot 10^{-3}$	-3.01	-3.18	+0.17
PMS	$1.5 \cdot 10^{-3}$	-3.85	-6.00	+2.15
PMAHA	$2.0 \cdot 10^{-8}$	-10.50	-8.52	-1.98
AHA+PMS	-	-6.86	-9.18	+2.35
$\Delta\Delta$(kcal/mol)	-	-3.64	+0.66	-4.30

Table 1: Corrected thermodynamic parameters for the investigated inhibitors.

6.4 Discussion

The large affinity observed for PMAHA is in line with expectations for this class of MMP inhibitors. The deconstructing analysis of this simple linked molecule provides a new piece of evidence to rationalize the key factors playing a role in the tethering approach.

Although theoretical models predict, for the dimerization of adjacent ligands, affinity advantages up to 10^{-8} M with respect to either fragment, the experience on real cases shows that usually the gain in affinity is orders of magnitude smaller[20,44]. In fact, the linkage of the fragments often results in a new molecule that does not allow each of them to adopt the original and optimal binding mode[25].

In this respect, PMAHA with the related fragments PMS and AHA constitutes a good case study because all the requirements relevant to maximize the linking effect are satisfied in the molecule. In particular, the structural data show that the binding mode of the two individual fragments is substantially maintained in the joined molecule. Moreover, no sizable additional favorable interactions or unfavorable strain energies are introduced by the linker since the two fragments are “tethered” by a single covalent bond without any additional intervening atom.

As already shown in table 1, PMAHA exhibits a dissociation constant (K_D^{AB}) of 2.0×10^{-8} M, much lower than the dissociation constants of the two fragments, PMS ($K_D^A = 1.5 \times 10^{-3}$ M) and AHA ($K_D^B = 6.2 \times 10^{-3}$ M), and nearly three order of magnitude lower than the product of their relative affinity constants.

A suitable approach to investigate the linking coefficient is to compare the difference in experimental ΔG^0 and ΔH^0 values and in the calculated entropy contributions. For PMAHA the difference in experimental ΔG^0 ($\Delta\Delta G^0$) with respect to the sum of the ΔG^0 values for the

isolated PMS and AHA is large and negative (-3.64 kcal/mol). This increased affinity is not related to an improved interaction, as shown by the small (and even positive) difference (+0.66 kcal/mol) in experimental ΔH^0 ($\Delta\Delta H^0$) between PMAHA and the isolated PMS and AHA. On the contrary, the difference in $T\Delta S^0$ for PMAHA is high and negative (-4.30 kcal/mol) and fully accounts for the large linking effect observed for the tethered molecule.

Besides the calculation of the value for the linking coefficient for the investigated molecule, the present analysis allows to us to achieve a more general view on the entropic contribution in fragment-based drug design. In particular, in the case of PMAHA no additional enthalpic and entropic terms, arising from atoms of the linker, contribute to ΔH^0 and to ΔS^0. At the same time, other possible contributions to ΔS^0 are probably negligible, as suggested by the superimposition of the complexes and by the small difference in the interaction energy ($\Delta\Delta H^0$ = -0.66 kcal/mol). Collectively, all these findings provide an estimation of the value for the linking coefficient that can be reasonably achieved by tethering two ligands when the interactions of each of them are maintained and no additive effects arise from the atoms of the linker.

The octanol/water partition coefficient (log *P*) calculated as CLog P value for each of the two fragments and for the tethered molecule shows that the hydrophilicity increases of one order of magnitude from PMS (+0.42) to PMAHA (-0.69).

In this respect, for PMAHA an increase in free energy of binding even larger than 4.30 kcal/mol would have been possible without the small unfavorable enthalpic term of +0.66 kcal/mol possibly related to small changes in binding mode. This large value opens also interesting questions on the advantages provided by the linking coefficient in fragment-linking and fragment growing strategies[45]. In the former, additional atoms are introduced to tether the fragments, without altering the original binding mode. Obviously the length and the structure of the linker is designed to avoid steric clashes, to reduce the degrees of freedom and, possibly, to establish interactions with the target. However the linker is not conceived to maximize the interactions these interactions, and detrimental entalpic and entropic contributions to the free energy of binding can drastically reduce the benefit of the linking coefficient[46]. On the contrary, in the fragment growing approach at each step the additional moiety introduced in the scaffold is conceived to establish new strong interactions with the target without altering the optimal binding mode of the original core. Therefore, the latter strategy might theoretically take the best advantage from the entropic gain related to the linking coefficient. However, it should be pointed out that, in real cases, several factors such

as the lack of detailed structural information and the protein flexibility make it difficult to obtain a suitable growing of the fragment, and the two strategies usually result in ligands with similar affinity constants[47].

In summary we have shown that when the two fragments are tethered with a zero-length linker in a way that all the main interactions are maintained, linking coefficient values approaching 10^{-3} M^{-1} can be obtained. We have also verified the entropic nature of the linking coefficient, that in the absence of unfavourable enthalpic contribution can be even smaller than 10^{-3} M^{-1}.

Reference List

1. Shuker,S.B., Hajduk,P.J., Meadows,R.P. & Fesik,S.W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531-1534 (1996).

2. Erlanson,D.A. *et al.* Site-directed ligand discovery. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 9367-9372 (2000).

3. Kolb,H.C., Finn,M.G. & Sharpless,K.B. Click chemistry: Diverse chemical function from a few good reactions. *Angewandte Chemie-International Edition* **40**, 2004-+ (2001).

4. Bajorath,F. Integration of virtual and high-throughput screening. *Nature Reviews Drug Discovery* **1**, 882-894 (2002).

5. Bleicher,K.H., Bohm,H.J., Muller,K. & Alanine,A.I. Hit and lead generation: Beyond high-throughput screening. *Nature Reviews Drug Discovery* **2**, 369-378 (2003).

6. Blundell,T.L., Jhoti,H. & Abell,C. High-throughput crystallography for lead discovery in drug design. *Nat. Rev. Drug Discov.* **1**, 45-54 (2002).

7. Johnson S.L. & Pellecchia M. Structure- and fragment-based approaches to protease inhibition. *Curr Top Med Chem* **6**, 317-329 (2006).

8. Congreve,M., Carr,R., Murray,C. & Jhoti,H. A 'rule of three' for fragment-based lead discovery? *Drug Discov. Today* **8**, 876-877 (2003).

9. Rix,U. & Superti-Furga,G. Target profiling of small molecules by chemical proteomics. *Nature Chemical Biology* **5**, 616-624 (2009).

10. Bertini,I., Fragai,M., Lee,Y.-M., Luchinat,C. & Terni,B. Paramagnetic metal ions in ligand screening: the Co^{II} matrix metalloproteinase 12. *Angew. Chem. Int. Ed.* **43**, 2254-2256 (2004).

11. Bertini,I., Fragai,M., Luchinat,C. & Talluri,E. Water-based ligand screening for paramagnetic metalloproteins. *Angew. Chem. Int. Ed.* **47**, 4533-4537 (2008).

12. Pellecchia,M. *et al.* Perspectives on NMR in drug discovery: a technique comes of age. *Nature Rev. Drug Discov.* **7**, 738-745 (2008).

13. Hajduk,P.J. & Greer,J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discov.* **6**, 211-219 (2007).

14. Albert,J.S. *et al.* An integrated approach to fragment-based lead generation: philosophy, strategy and case studies from AstraZeneca's drug discovery programmes. *Curr. Top. Med. Chem.* **7**, 1600-1629 (2007).

15. Chen,Y. & Shoichet,B.K. Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* **5**, 358-364 (2009).

16. Ciulli,A., Williams,G., Smith,A.G., Blundell,T.L. & Abell,C. Probing hot spots at protein-ligand binding sites: A fragment-based approach using biophysical methods. *Journal of Medicinal Chemistry* **49**, 4992-5000 (2006).

17. Pellecchia,M. Fragment-based drug discovery takes a virtual turn. *Nat. Chem. Biol.* **5** , 274-275 (2009).

18. Hajduk,P.J. *et al.* Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. *J. Am. Chem. Soc.* **119**, 5818-5827 (1997).

19. Olejniczak,E.T. *et al.* Stromelysin inhibitors designed from weakly bound fragments: Effects of linking and cooperativity. *J. Am. Chem. Soc.* **119**, 5828-5832 (1997).

20. Page,M.I. & Jencks,W.P. Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc. Natl. Acad. Sci. U. S. A* **68**, 1678-1683 (1971).

21. Jencks,W.P. On the attribution and additivity of binding energies. *Proc. Natl. Acad. Sci. USA* **78**, 4046-4050 (1981).

22. Murray,C.W. & Verdonk,M.L. The consequences of translational and rotational entropy lost by small molecules on binding to proteins. *Journal of Computer-Aided Molecular Design* **16**, 741-753 (2002).

23. Pellecchia,M. Solution nuclear magnetic resonance spectroscopy techniques for probing intermolecular interactions. *Chem. Biol.* **12**, 961-971 (2005).

24. Congreve,M.S. *et al.* Detection of ligands from a dynamic combinatorial library by X-ray crystallography. *Angew. Chem. Int. Ed Engl.* **42**, 4479-4482 (2003).

25. Babaoglu,K. & Shoichet,B.K. Deconstructing fragment-based inhibitor discovery. *Nature Chemical Biology* **2**, 720-723 (2006).

26. Babine,R.E. & Bender,S.L. Molecular Recognition of Protein-Ligand Complexes: Applications to Drug Design. *Chem. Rev.* **97**, 1359-1472 (1997).

27. Pavlaki,M. & Zucker,S. Matrix metalloproteinase inhibitors (MMPIs): the beginning of phase I or the termination of phase III clinical trials. *Cancer Metastasis Rev* **22**, 177-203 (2003).

28. MacPherson,L.J. *et al.* Discovery of CGS 27023A, a non-peptidic, potent, and orally active stromelysin inhibitor that blocks cartilage degradation in rabbits. *J. Med. Chem.* **40**, 2525-2532 (1997).

29. Aureli,L. *et al.* Structural bases for substrate and inhibitor recognition by matrix metalloproteinases. *Curr. Med. Chem.* **15**, 2192-2222 (2008).

30. Bertini,I., Fragai,M. & Luchinat,C. Intra- and Interdomain Flexibility in Matrix Metalloproteinases: Functional Aspects and Drug Design. *Curr. Pharm. Des.* **15**, 3592-3605 (2009).

31. Bertini,I. *et al.* Crystal structure of the catalytic domain of human matrix

metyalloproteinase 10. *J. Mol. Biol.* **336**, 707-716 (2004).

32. Bertini,I. *et al.* Conformational variability of MMPs: beyond a single 3D structure. *Proc. Natl. Acad. Sci. USA* **102**, 5334-5339 (2005).

33. Bertini,I. *et al.* Exploring the subtleties of drug-receptor interactions: the case of matrix metalloproteinases. *J. Am. Chem. Soc.* **129**, 2466-2475 (2007).

34. Leslie,A.G.W. Molecular data processing. Moras,D., Podjarny,A.D. & Thierry,J.-C. (eds.), pp. 50-61 (Oxford University Press, Oxford,1991).

35. Evans,P.R. "Data Reduction", Proceedings of CCP4 Study Weekend. Data Collection & Processing,114-122.1993. Ref Type: Conference Proceeding

36. Rossmann,M.G. & Blow,D.M. The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.* **D15**, 24-31 (1962).

37. Crowther,R.A. Rossmann,M.G. (ed.) (Gordon & Breach, New York,1972).

38. Vagin,A. & Teplyakov,A. An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr* **56**, 1622-1624 (2000).

39. Vagin,A. & Teplyakov,A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022-1025 (1997).

40. Murshudov,G.N., Vagin,A.A. & Dodson,E.J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst.* **D53**, 240-255 (1997).

41. McRee,D.E. "XtalView: A Visual Protein Crystallographic Software System for XII/XView". *J. Mol. Graphics* **10**, 44-47 (1992).

42. Lamzin,V.S. & Wilson,K.S. Automated refinement of protein models. *Acta Crystallogr. D. Biol. Crystallogr.* **49**, 129-147 (1993).

43. Laskowski,R.A., MacArthur,M.W., Moss,D.S. & Thornton,J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291 (1993).

44. Whitty,A. Cooperativity and biological complexity. *Nat. Chem. Biol.* **4**, 435-439 (2008).

45. Ciulli,A. & Abell,C. Fragment-based approaches to enzyme inhibition. *Curr. Opin. Biotechnol.* **18**, 489-496 (2007).

46. Chung,S., Parker,J.B., Bianchet,M., Amzel,L.M. & Stivers,J.T. Impact of linker strain and flexibility in the design of a fragment-based inhibitor. *Nat. Chem. Biol.* **5**, 407-413 (2009).

47. Hung,A.W. *et al.* Application of Fragment Growing and Fragment Linking to the Discovery of Inhibitors of Mycobacterium tuberculosis Pantothenate Synthetase. *Angew. Chem. Int. Ed Engl.* (2009).

Chapter 7

Final conclusions and perspectives

The comprehension of the physiological role of proteins in biological systems requires an integrated study where the structural and dynamical analysis is flanked by a detailed investigation of the networks of the interactions involving each protein. Breaking and formation of weak interactions not only controls the protein folding and dynamics, but are crucial for the protein function as responsible for the formation of stable or transient protein-protein complexes. Weak interactions are also responsible for the binding of small molecules as substrates, cofactors or drugs to proteins. In this PhD research the role of weak interactions has been investigated in biologically relevant models using different biophysical and structural methodologies in order to clarify their contribution to protein folding, dynamics and to protein-protein and protein-ligand recognition. The research has been mainly focused on cloning, expression and characterization of human calcium binding proteins calmodulin, S100A16, S100P and on two soluble extracellular domains of the pathologically relevant RAGE receptor. In particular, samples of calmodulin have been exploited to investigate by relaxometry, weak and transient interdomain interactions and their effects on protein dynamics. The relaxometric analysis performed on apo calmodulin, in fact allowed us to monitor its behaviour in solution where, as shown by high field NMR measurements, the N-terminal and C-terminal domain experience a dynamic ensemble of conformations. The analysis of the collective relaxation rate of the non-exchangeable protein protons provided a reorientation time in agreement with the overall rotational motion of the protein when the two domains are arranged in the closed form. This indicates that the two domains preferentially assume several compact structures, with the two domains close to one another. This latter evidence and the absence of interdomain NOEs indicates that the inter-domain conformational heterogeneity must be the result of a number of slowly interconverting distinct conformations associated with the breaking and formation of weak interactions. Moreover the analysis suggested the presence of a large side chain mobility which could not be monitored through standard high field ^{15}N-^{1}H relaxation measurements. The work focused on S100A16 allowed us to express the WT protein in good yield and to develop a reliable protocol to produce selenium-enriched S100A16 in order to solve the problem of phase in X-ray crystallography experiments. Crystals of Se-S100A16 have been obtained in large amount but, unfortunately, the observed gemination prevented the collection of data suitable for structure resolution. On the contrary, the NMR analysis performed on apo- and holo-S100A16 samples has provided the first structural and dynamical characterization of this protein. As all the other S100 proteins, both apo and holo S100A16 in solution exist as homodimers stabilized by a network

of weak interactions. The structural analysis showed that the interaction of the calcium ion with the aminoacids forming the metal binding site induces a structural reorganization with exposition of hydrophobic surfaces. ^{15}N relaxation measurements and the RMSD per residues for the calculated families of structures showed that the hinge-loop region of the protein experiences a sizably large mobility.

Furthermore, the role of weak interactions in protein-protein recognition has been investigated by analyzing the binding of S100P to extracellular domains of RAGE. Identifying the interaction surface between the V-domain of RAGE and holo S100P provides an opportunity to understand the structural details of the interactions of RAGE receptor with one of its pathologically relevant ligands. Actually, RAGE interacts with a wide range of ligands that structurally have very little in common leading to diverse cellular responses from cytokine secretion and increased cellular oxidant stress to cell survival, differentiation, and proliferation. However, RAGE is a membrane receptor of and can not be investigated in solution by NMR. The research activity carried out on RAGE during this PhD has been focused on cloning, expression and characterization of two extracellular domains of RAGE, V and C12 domains and of its ligand S100P. We observed that the C1 domain of RAGE is not required for the binding to S100P because the pattern of shifts on holo S100P is the same upon the addition of V-domain and VC1 tandem construct. The NMR analysis clearly showed that one V-domain binds to one S100 homodimer. Using the chemical shifts perturbations observed on the S100P and on V-domain, a model of the complexes S100P-V-domain and also the related adduct with VC1 tandem construct have been calculated using Haddock. Understanding of the structural details of the interaction between RAGE and its ligands can provide new therapeutic opportunities to treat diseases as chronic inflammations, cancer and diabetes. Therefore, cloning, expression and structural characterization of the whole RAGE receptor and the analysis of its interaction with all the pathologically relevant ligands is the obvious development of the research started with this PhD.

In the last part of my PhD, the role of weak interactions in ligand binding and the thermodynamic aspects of tethering have been analyzed by performing a structural and calorimetric analysis on a aryl-sulfonamide inhibitor of MMP-12 and on its constituting fragments. The analysis performed on samples of catalytic domains of MMP-12 showed that the tethered molecule displays a very small value of the linking coefficient, entirely due to a very favorable entropic contribution with the enthalpic term being very similar to the sum of those of the two isolated fragments. Moreover, we proved that large contribution to the

binding energy can be achieved when the two fragments are tethered with a zero-length linker in a way that all the main interactions are maintained.

List of pubblications

Borsi, V., Luchinat, C., and Parigi, G. “Global and local mobility of apocalmodulin monitored through fast field cycling relaxometry” *Biophys J.*, **97**(6), 1765-61, 2009.

Borsi, V., Calderone, V. , Fragai, M., Luchinat, C. and Sarti N. “Entropic contribution to the linking coefficient in Fragment Based Drug Design: a case study.”, 2009 **Submitted**

“Structural basis of RAGE receptor activation by S100P.”
In preparation.

“Solution structure and dynamics of S100A16 in the apo and Ca^{2+} -bound states.”
In preparation.

J Biol Inorg Chem (2011) 16:243–256
DOI 10.1007/s00775-010-0721-3

ORIGINAL PAPER

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

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

Received: 16 July 2010 / Accepted: 9 October 2010 / Published online: 3 November 2010

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

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

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

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

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

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

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

Introduction

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

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

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

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

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

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

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

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

Materials and methods

Protein expression

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

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

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

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

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

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

Crystallization, data collection, and structure determination

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Isothermal titration calorimetry

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

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

NMR spectroscopy and solution structure determination

All NMR experiments for assignments were performed at 25 °C with a Bruker 500 MHz spectrometer equipped with a cryoprobe. Apo and calcium(II)-loaded S100A16 samples (0.6 and 0.8 mM, respectively) were ^{13}C,^{15}N-labeled, in 20 mM MES, 200 mM KCl, and 1 mM DTT buffer (pH 5.5), containing 10% D_2O. Sequential assignments of the backbone resonance were achieved via HNCO, HNCA, CBCA(CO)NH and HNCACB spectra. Side chain assignments were performed through 3D (H)CCH total correlation spectroscopy, HBHA(CBCACO)HN together with ^{13}C nuclear Overhauser effect spectroscopy (NOESY) heteronuclear single quantum coherence (HSQC) and ^{15}N-NOESY HSQC experiments. Proton–proton distance restraints were derived from the analysis of 2D-NOESY, ^{15}N-NOESY-HSQC, and ^{13}C-NOESY-HSQC spectra acquired with a Bruker 900 MHz spectrometer equipped with a cryoprobe. The spectra were processed using TOPSPIN 2.0 and analyzed with CARA [27]. Backbone dihedral angles were obtained from TALOS+ [28] from the chemical shifts of N, HN, H^{α}, C, C^{α}, and C^{β} nuclei. The structures were calculated using the program CYANA-2.1 [29, 30] by imposing the dimer symmetry constraint (noncrystallographic symmetry constraint). The two subunits in the dimeric structure were linked together through a chain of dummy atoms with zero van der Waals radii. The calcium(II) ions were included in the calculation of the calcium-loaded form by adding new residues in the amino acid sequence. Four chains of dummy atoms with zero van der Waals radii, which can freely penetrate into the protein, each of them ending with one atom with a radius of 1.8 Å, which mimics the calcium ion, were included for this purpose. Protein ligand atoms were linked to the metal ion through upper distance limits of 3 Å, according to the structure of S100A13.

The best 30 structures out of the calculated 350 structures of the CYANA family were then subjected to restrained energy minimization with AMBER 10 [31]. Nuclear Overhauser effect (NOE) and torsion angle restraints were applied with force constants of 50 kcal mol^{-1} $Å^{-2}$ and 32 kcal mol^{-1} rad^{-2}, respectively. The programs PROCHECK-NMR [32] and WHATIF [33] were used to evaluate the quality of the structures.

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

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

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

Heteronuclear relaxation measurements

^{15}N-R_1, R_2, and steady-state heteronuclear ^{1}H–^{15}N NOEs were measured using a 700 MHz spectrometer using standard pulse sequences [34, 35], at 25 °C. The longitudinal (R_1) and transverse (R_2) relaxation rates were determined by fitting the cross-peak intensities as a function of the delay to a single-exponential decay through the standard routines of the Sparky program [36]. The heteronuclear NOE values were obtained from the ratio of the peak height for ^{1}H-saturated and unsaturated spectra. The heteronuclear NOE values and their errors were estimated by calculating the mean ratio and the standard error from the available data sets. R_1, R_2, and NOE values were obtained for 91 out of the 102 assigned backbone ^{N}H resonances for both the apo and the calcium forms. Estimates of the reorientation time were then calculated with the model-free approach [37] and S^2 values were calculated with the program TENSOR2 [38]. Theoretical predictions of ^{N}H R_1 and R_2 values for apo-S100A16 and calcium(II)-loaded S100A16 were calculated using HYDRONMR [39].

Results

Isothermal titration calorimetry

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

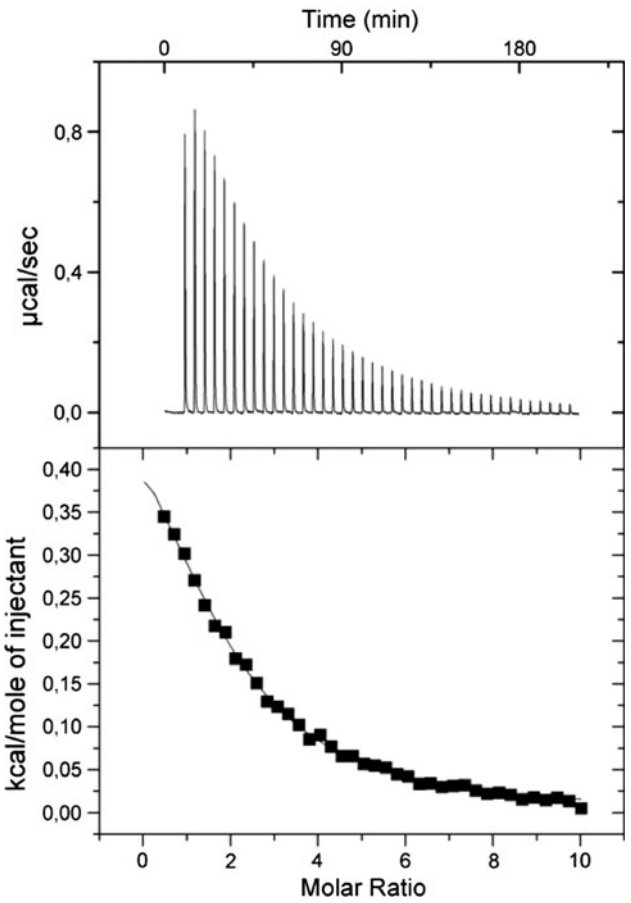


Fig. 1 Isothermograms for the binding of S100A16 to Ca^{2+}. The raw data and the fit to the one binding site model are reported in the *upper panel* and the *bottom panel*, respectively. The fit performed using a sequential two binding sites model is of similar quality. Appropriate background corrections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

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

NMR resonance assignments

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

Ca^{2+} titration of apo-S100A16

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

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

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

^{15}N relaxation measurements

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

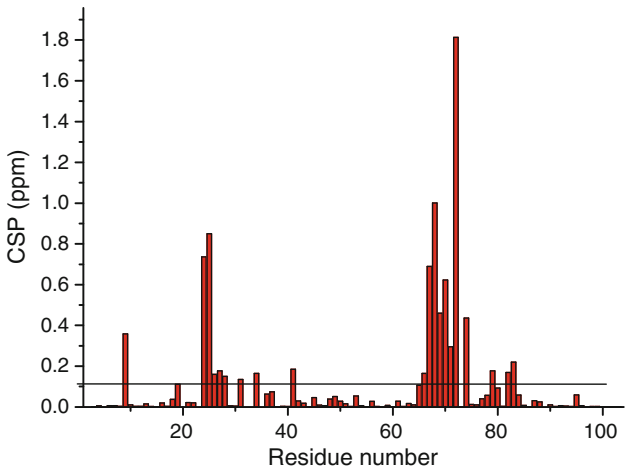


Fig. 2 Chemical shift perturbation (*CPS*) of S100A16 upon calcium(II) binding. The *horizontal line* indicates the average value. Shift perturbations are reported as a weighted average of the amide proton and amide nitrogen shifts using the formula $\Delta\delta = [(\Delta\delta_H)^2 + (\Delta\delta_N/5)^2]^{0.5}$

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

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

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

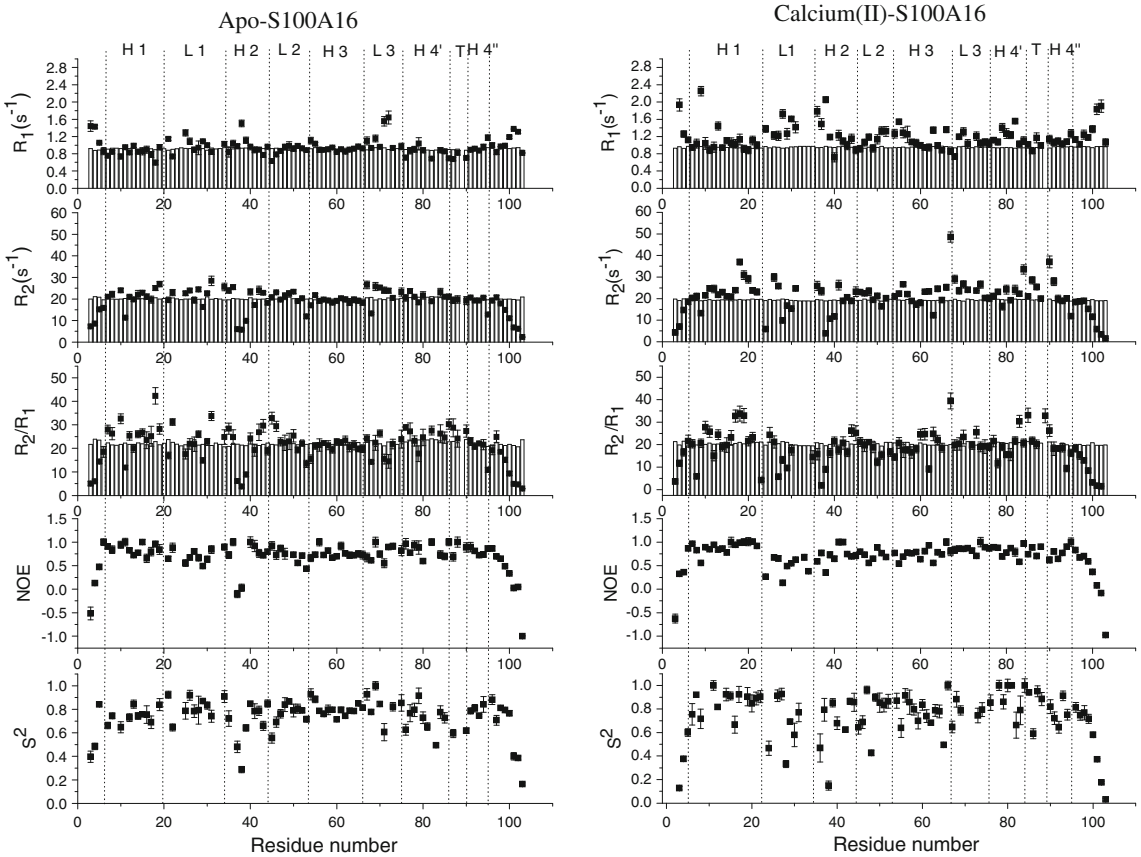


Fig. 3 Sequential plot of the experimental relaxation parameters of apo-S100A16 and calcium(II)-loaded S100A16. The values calculated with HYDRONMR are shown as *bars*

Solution structure of apo-S100A16 and calcium-loaded S100A16

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

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

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

Crystal structure of apo-S100A16

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

Table 2 Structural restraints and statistical analysis

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

NOE nuclear Overhasuer effect, *RMS* root mean square

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

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

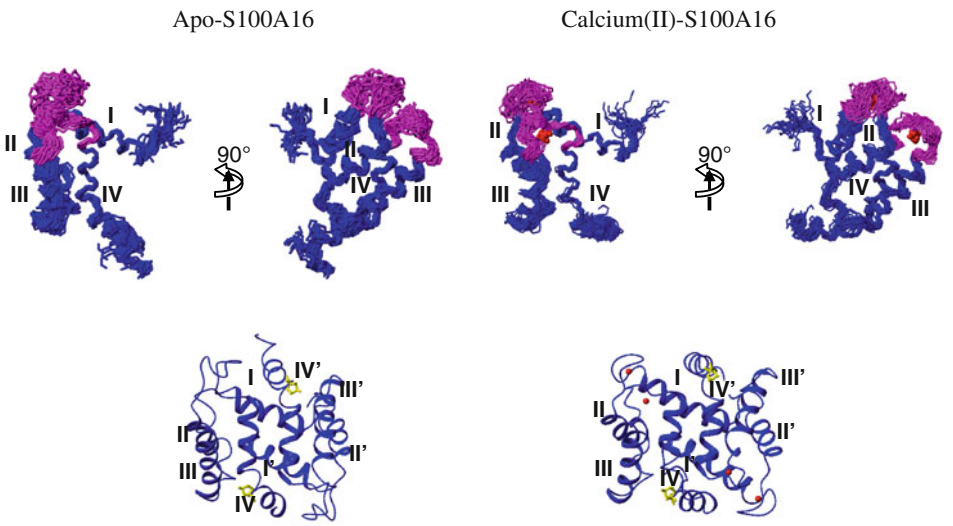


Fig. 4 Solution structural families of the S100A16 subunit and ribbon representation of the homodimer mean structures of the protein in the apo and calcium(II)-bound forms. *Roman numerals* indicate the helix numbers. Calcium ligand loops are shown in *magenta*. Pro-89 is shown in *yellow*

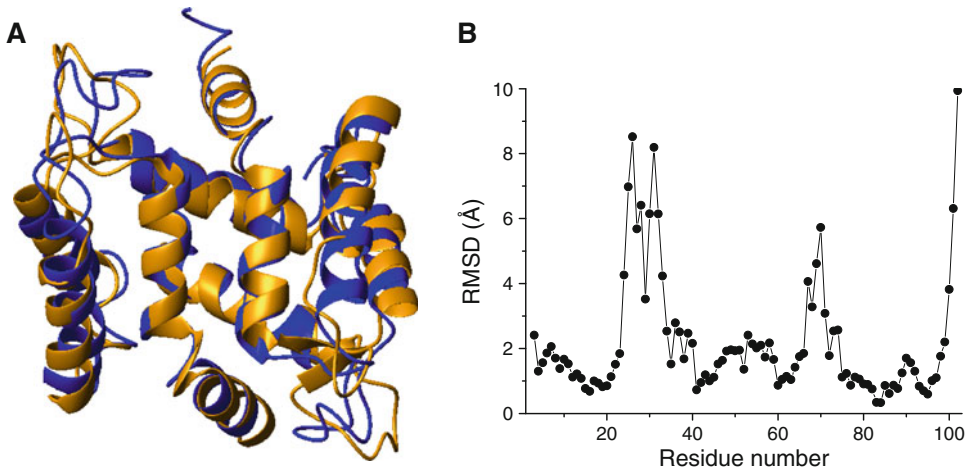


Fig. 5 **a** Three-dimensional structure of one of the two homodimers in the crystal structure (in *yellow*) superimposed on the mean solution NMR structure (in *blue*) of apo-S100A16, and **b** backbone root mean square deviation (*RMSD*) between the crystal structure and the mean average structure

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

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

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

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

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

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

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

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

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

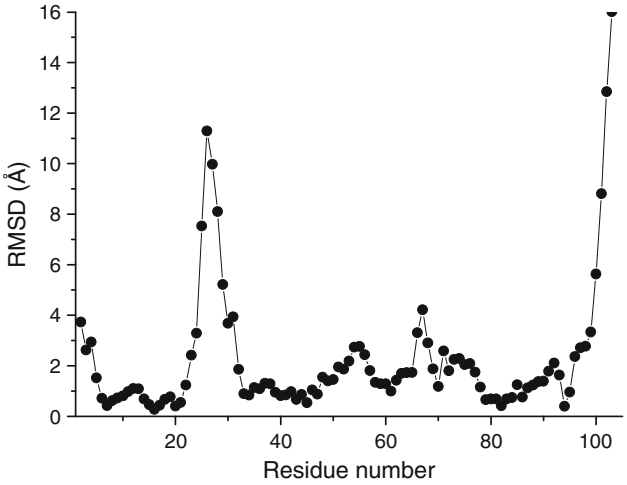


Fig. 6 Backbone RMSD between the apo-S100A16 and the calcium(II)-loaded S100A16 mean solution structures

Discussion

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

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

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

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

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

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

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

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

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

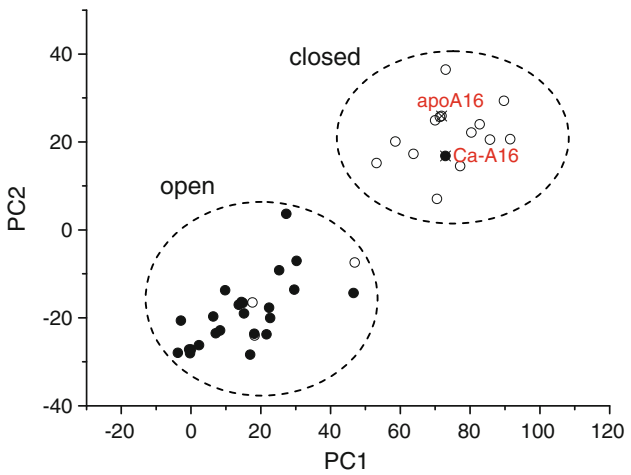


Fig. 7 Principal component plot for the S100 proteins derived from principal component analysis of the six interhelical angles. Apoproteins (S100A1, S100A2, S100A3, S100A4, S100A5, S100A6, S100A10, S100A11, S100A13, S100A16, S100B, calbindin D_{9k}) are indicated with *open circles* and calcium(II)-bound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

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

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

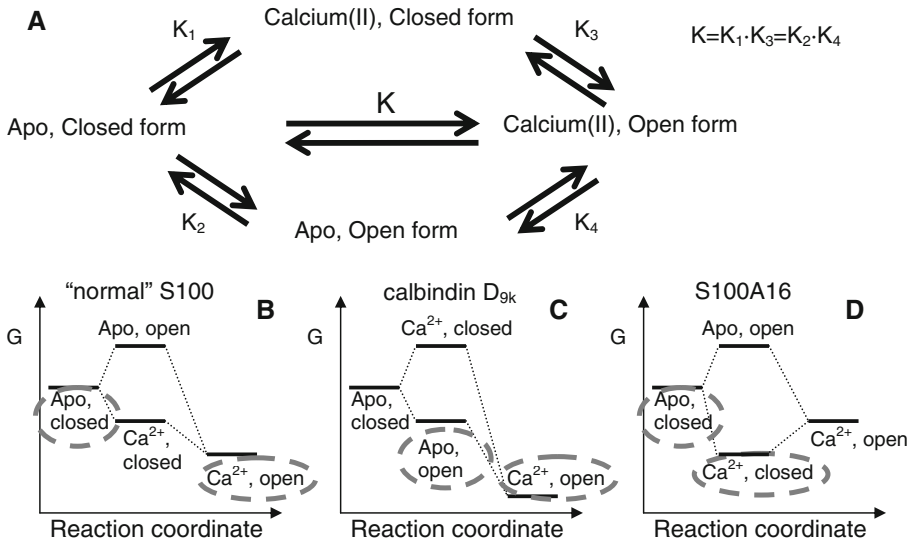


Fig. 8 Equilibrium constants (**a**) and energy levels (**b**–**d**) for the of open and closed forms of the S100 structures in the apo and calcium(II)-bound states. Observable forms of "normal" S100 proteins (**b**), calbindin D_{9k} (**c**), and S100A16 (**d**) are highlighted with *dashed lines*

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

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

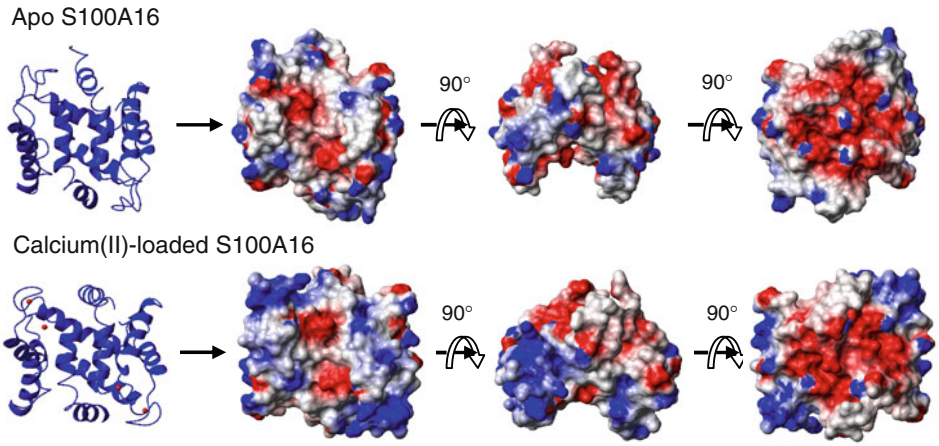


Fig. 9 Electrostatic surface representation of the S100A16 dimer

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

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

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

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

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

References

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

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