University of Florence

Doctorate in Structural Biology

Cycle XXI (2006-2008)

STUDY OF PROTEIN-PROTEIN INTERACTIONS AND THEIR FUNCTIONAL IMPLICATIONS

Ph. D. thesis of

Nesi Antonella

Tutor and 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.

Alla mia famiglia …

"La felicità
non è avere quello che si desidera,
ma desiderare quello che si ha"
(Oscar Wilde)

Contents

I. INTRODUCTION

I.1 STRUCTURAL BIOLOGY

Proteins are one of the four main classes of molecules, along with carbohydrates, fats, and nucleic acids, that underlie all life, since they are responsible of most cellular function. The massive advances in genomics research over the past few years have now led to a renewed focus on protein structure and function.

Genome sequencing projects have provided a list of proteins contained in the cell. This list is still incomplete, as it does not always capture variants like alternative spliced forms or post-translational modifications. Nevertheless, it provides the scaffold onto which most functions lie.

The successes of high-throughput approach in genomics (DNA sequencing, DNA microarrays) have inspired similar initiatives in protein science, with high-throughput programmes for 3D structure determination[1].

Structural biology has emerged as a powerful approach for defining the functions of proteins; this capacity is based on the observation that the evolutionary constraints for three-dimensional structures of proteins are higher than for sequences[2,3].There are many cases of distantly related homologues assignable from shared structures with no recognizable relationship between their sequences. Many algorithms have been implemented for alignment by structural analogy.

The strong predictive power of structure in functional annotation has resulted in the rapid growth of the new field of structural genomics (SG) (or structural proteomics)[4] and to the rapid development of novel high-throughput technologies[4,5]. In addition to expediting functional characterization of gene products, SG initiatives also provide a comprehensive view of the protein structure universe, by determining the structures of representative proteins from every protein fold family[6]. In the same time, bioinformatics collects data on sequences, structures, and functions, and studies the correspondences between them.

Over the past 10 years, several international structural genomics initiatives have been funded with diverse approaches[7-10]. For instance, the Protein Structure Initiative (PSI) in the United

States, has been devoted to the study of the complex relationships between the evolution of the function of proteins with respect to the evolution of their sequence and structure.

Both the number of sequences and protein families are still growing at an exponential rate, but domain families are 10-fold fewer (<10 000) than the number of protein families, for this reason the structural genomics initiatives consider domains as the fundamental unit of both protein structure and evolution. For instance, the SCOP database has as its basis individual domains of proteins. Sets of domains are grouped into families of homologues, for which the similarities in structure, sequence, and sometimes function, imply a common evolutionary origin [http://scop.mrc-lmb.cam.ac.uk/scop /index.html].

Like most sequence-based methods, these structure-based methods proceed by searching for homologues and do not permit unambiguous assignment of a precise function: even closely related proteins can have different functions[11], conversely, non-homologous proteins may have similar functions, moreover there are numerous examples of proteins with multiple functions.

For instance, multi-domain proteins present particular problems for functional annotation, because each domain may possess independent functions, modulate one another's function, or act in concert to provide a single function.

For these limits of structure-based methods, a functional genomic approach, making use of contextual information and intergenomic comparisons, is useful to predict a protein function on the basis of inferences from genomic contexts and protein interaction patterns[12-15]. Functional genomics can use different approaches for predictions of protein functions, deriving information from relationship between non-homologous proteins[16]:

– Gene fusion. A composite gene in one genome may correspond to separate genes in other genomes. For instance, the proportion of multidomains protein is higher in eukaryota with respect to prokaryota[17].

– Local gene context. Analyse co-regulated and co-transcribed components of a pathway. (In bacteria, genes in a single operon are usually functionally linked).

– Interaction patterns. The network of interactions reveals the function of a protein.

– Phylogenetic profiles. Proteins in a common structural complex or pathway are functionally linked and expected to co-evolve[18].

Comparison of the flexibilities of homologous proteins across species suggested that, as the species gets more complex, its proteins become more flexible. In fact the number of genes in

the human genome is fewer than that of some lower organisms but our genome is more flexible and functionally more complex[19].

In contrast to the classical view of structured proteins, the concept of intrinsically disordered regions has recently emerged[20-25]. Disordered regions are protein segments that does not completely fold and remains flexible and unordered, existing in a continuum of conformations from the less to the more structured states[26].

The importance of intrinsically disordered proteins rely on their involvement in a broad range of functions[27]. Moreover, knowledge of the folded and denatured states under different conditions, can help in the comprehension of protein folding. Genomic analysis of disordered proteins indicates that the proportion of the genome encoding intrinsically unstructured proteins increases with the complexity of organisms[28]: computational predictions estimate that proteomes of archaea and bacteria comprise only a small fraction of intrinsically disordered proteins (about 2–4%), while eukaryotic proteomes include a large fraction (about 33%) of long regions that are natively disordered and thus do not adopt a fixed structure[29]. Disordered regions of proteins have been shown to have key physiological roles, for example, are involved as communicators in many cellular signalling pathways[30-32].

Disordered functional proteins provide evidence that the function of a protein and its properties are not only decided by its static folded three-dimensional structure; they are determined by the distribution and redistribution of the conformational substates[33].

Intrinsically disordered proteins can be broadly classified into two major groups: those that are fully disordered throughout their length (often called natively unfolded proteins) and those that have extensive (>30–40 residues) regions that are disordered and embedded in an otherwise folded protein. Natively unfolded proteins can be further subdivided into two groups, those with no ordered secondary structure and those with some secondary structure; the latter resemble molten globules and lack tertiary structure.

One distinction in the amino acid sequences of natively unfolded proteins has been suggested in the literature, like the presence of numerous uncompensated charged groups (often negative) at neutral pH, arising from the extreme pI values in such proteins. A low content of hydrophobic amino acid residues has been also noted for several natively unfolded proteins. Moreover, disordered regions of proteins are characterized by low sequence complexity, high flexibility and amino acid compositional bias: compared to sequences of ordered proteins, intrinsically disordered segments and proteins have significantly higher levels of certain amino acids (E, K, R, G, Q, S and P) and lower levels of others (I, L, V, W, F, Y, C and N). For these features, long disordered polypeptide sequences can be predicted successfully from

amino acid sequence[34]. For instance, FoldIndex is a program that estimates the local and general probability of the provided sequence to fold [http://bip.weizmann.ac.il/fldbin/findex].
Recent studies have identified a natively folded protein that break the Anfinsen's hypothesis: "a polypeptide achieves its biologically active, native state by descending to the most thermodynamically stable configuration, which corresponds to one of a few thousand unique folds, with varying amounts of local flexibility". The human chemokine lymphotactin (Ltn) adopts two distinct folds at equilibrium in physiological conditions, and interconversion between the conformers involves almost complete restructuring of its hydrogen bond network and other stabilizing interactions, in contrast to other cases of different conformers that share a large common substructure during interconversion[35,36].
Because each Ltn conformer displays only one of the two functional properties essential for its activity *in vivo* (Receptor activation and Glycosaminoglycans binding), the conformational equilibrium is likely to be essential for the biological activity of lymphotactin and could represent a novel regulatory mechanism for proteins functions.
Therefore, it is clear that proteins with large unstructured regions and natively unfolded proteins have a very important physiologic role.
This research project has been focused on the expression of proteins with large unstructured regions, which are involved in protein-protein interactions and in cell signalling.

I.2. PROTEIN-PROTEIN INTERACTIONS

A practical way to understand protein functions is the identification of binding partners. Valuable information on cellular pathways can be obtained by investigating protein-protein interaction. These data are also helpful in drug design and to evaluate the role of mutations, which are often clustered in binding sites[37].
The recent increase in the number of protein structures, the additional experimental results of protein-protein interactions indicate that some proteins are centrally connected, whereas others are at the edges of the map[38].
The centrally connected proteins may interact with a large number of partners and usually act as linkers of cellular processes, as regulatory elements in the organization of higher order protein interactions networks. Such proteins are usually those that perform the same function for many partners (phosphatases, kinases, transporters,…). The interface of such proteins

preferentially consists of α-helices. The capability to interact with a broad range of partner proteins is related to their higher content of repeat domains. These domains are easy to make (by duplication) and offer an opportunity to divergently evolve. At the same time, there is evidence that proteins whose function requires more specific interactions evolve slowly.

A large fraction of cellular proteins that play roles in cell-cycle control, signal transduction, transcriptional and translational regulation, and large macromolecular complexes, are estimated to be natively disordered[39]. Their native conformation can be stabilized upon binding. The global fold of disordered proteins does not change upon binding to different partners; however, local conformational variability can be observed, complicating the predictions of protein interactions. Upon binding, the equilibrium shifts in favour of the complex formation, further driving the reaction. As binding and folding are similar processes with similar underlying principles, this principle applies to disordered molecules in binding and to unstable, conformationally fluctuating building blocks in folding.

Protein-protein interactions are largely driven by the hydrophobic effect however hydrogen bonds, electrostatic interactions, and covalent bonds are also important[40].

The Gibbs free energy upon complex formation (binding free energy) can be determined directly from the equilibrium constant of the reaction (usually denoted as K_a and K_d, for association or dissociation constants). The equilibrium constants is function of the concentrations of both the free proteins and the complex at thermodynamic equilibrium.

The range of K_d values observed in biologically relevant processes is extremely wide and can span over twelve orders of magnitude. Weak protein-protein interactions, especially those with $K_d > 10^{-4}$ M, have been poorly characterized, despite they might be crucial for mediating many important cellular events[41].

An enormous number of enzymes, carrier proteins, scaffolding proteins, transcriptional regulatory factors, etc. function as oligomers.

Oligomerization and function in oligomeric proteins can be very finely tuned by ligand concentration (including ions, substrate, allosteric ligands, protons, etc.) and by protein concentration, influenced by expression levels, transport mechanisms or degradation rates. The formation of transient protein–protein complexes depends on the functional state of the partners. The affinities of such complexes are modulated at different levels, including interaction with ligands, other proteins, nucleic acids, ions such as Ca^{2+}, and covalent modification, such as specific phosphorlyation or acetylation reactions.

The interfaces in transient complexes are generally less extensive and more polar or charged, moreover the surfaces of interacting proteins at their interface are not optimized, leading to weaker associations, with the exception of some enzyme-inhibitor complexes[42].

On the contraty, obligatory complexes are in general tighter, with a stronger hydrophobic effect, better packing and fewer structural water molecules between the monomers, and they have better shape complementarity[43].

The characterization of structural motifs and domains involved in protein-protein interactions is important to understand the networks relevant for living cell and possible roles in diseases[44].

I.3. SYSTEM BIOLOGY

Molecular biology has until now mainly focussed on individual molecules, on their properties as isolated entities or as complexes in very simple model systems.

Scientific and technical advances triggered an exponential increase in the number of 3D structures deposited in the Protein Data Bank[45,46]. The number of protein complexes equally increased, providing thousands of different templates to model protein–protein interfaces.

Biological molecules in living systems participate in very complex networks, including regulatory networks for gene expression, intracellular metabolic networks and both intra- and intercellular communication networks. Such networks are involved in the maintenance (homeostasis) as well as the differentiation of cellular systems of which we have a very incomplete understanding.

Complete genomes, interaction and functional data must be integrated with 3D structures to build large cellular systems from their individual molecular components in order to understand how complex systems function and evolve. That is, scaling up from molecular biology to systems biology.

The aim of systems biology is the quantitative analysis and reconstruction of the structure and dynamics of cellular pathways via an in silico representation of the studied pathways, by an iterative process of matching experimental observations against model predictions to formulate new models and new experiments to test them[47,47,48].

In particular, it is important to define all of the components of the system, including the regulatory relationships between genes and interactions of proteins and biochemical pathways, and to use this knowledge to formulate a primitive model (biochemical or

mathematical) [49]. Once the system structure is defined, system behaviour can be analysed further using specific genetic and/or environmental perturbations. The data generated from such an analysis can either be integrated with the initial model or used to refine the model, such that its predictions are consistent with the experimental observations. The importance of systems biology research rely on the identification of novel protein functions or partners and the understanding of mechanisms that control the state of the cells so that they can be used to identify potential therapeutic targets for treatment of diseases.

I.4. MOLECULAR BIOLOGY

Understanding the molecular mechanisms by which cells carry out signal transduction pathways and control cell homeostasis, is an important goal to increase our knowledge of biology and to develop specific therapies for diseases. In this thesis will be discussed proteins, interactions and mechanisms involving proteins responsible for intra- and extracellular signaling pathways.

Our work has been focused on a group of proteins involved in intra- and extra-cellular signalling pathways.

I.4.1. Matrix Metalloproteinases

Matrix metalloproteinases represent a large family of 23 zinc-dependent endopeptidases in human, that on the basis of substrate specificity, sequence similarity, and domain organization, can be divided into six groups: Collagenases, Gelatinases, Stromelysins, Matrilysins, Membrane-Type MMPs, Other MMPs.

The MMPs share common structural and functional elements (Fig. 1).

Most members of the MMP family are secreted and organized into three well-conserved domains: an aminoterminal propeptide; a catalytic domain; and a hemopexin-like domain at the carboxy-terminal[50]. The propeptide consists of approximately 80–90 amino acids containing a cysteine residue, which interacts with the catalytic zinc atom via its side chain thiol group. Removal of the propeptide by proteolysis results in zymogen activation, as all members of the MMP family are produced in a latent form. The catalytic domain contains two

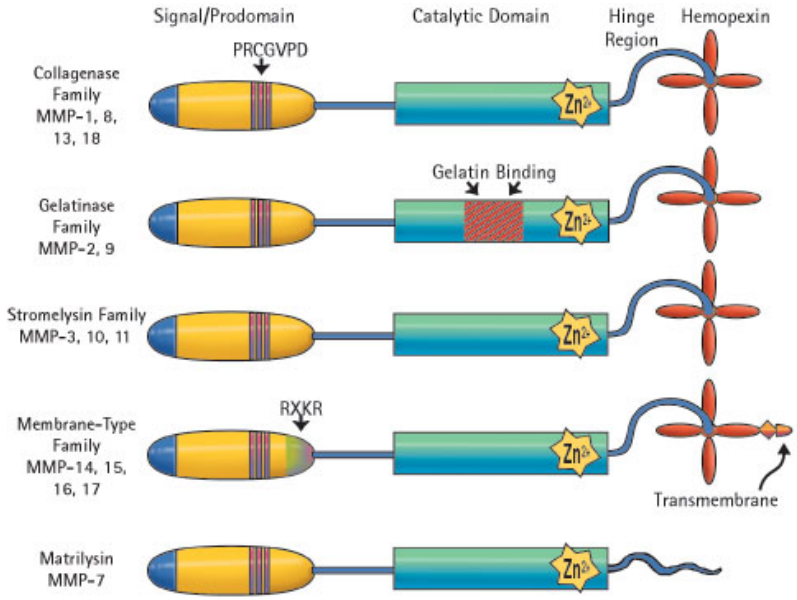


Fig.1. Domains organization in different classes of MMPs.

zinc ions and at least one calcium ion. One of the two zinc ions is structural, the other is essential for the proteolytic activity of MMPs and the three histidine residues that coordinate the catalytic zinc are conserved among all the MMPs. The hemopexin-like domain of MMPs is highly conserved and shows sequence similarity to the plasma protein hemopexin. The hemopexin-like domain has been shown to play a functional role in substrate binding and/or in interactions with the tissue inhibitors of metalloproteinases (TIMPs), a family of specific MMP protein inhibitors. In addition to these basic domains, the family of MMPs evolved into different subgroups by incorporating and/or deleting structural and functional domains. For instance, fibronectin type-II like domain in the gelatinases, transmembrane domain at the carboxy terminus and recognition motif (RXKR) for furin-like convertases at the end of the propeptide domain are characteristics of the membrane-type MMPs (MT-MMPs)[51,52].

MMPs are best known for their functions in remodeling of extracellular matrix and for their important roles in wound healing, angiogenesis, and invasive properties of cancer cells. Extracellular matrix (ECM) macromolecules are important for creating the cellular environments required during development and morphogenesis. Modulation of cell–matrix interactions and the integrity and composition of the ECM structure regulate cell proliferation, differentiation, and cell death.

Since MMPs are the major group of enzymes that regulate cell–matrix composition, loss of MMPs activity control may result in diseases such as arthritis, cancer, atherosclerosis, aneurysms, nephritis, tissue ulcers, and fibrosis. Moreover, cell membrane proteins have also

been identified as MMP substrates, expanding the potential importance of this family to include direct effects on cell-cell signalling, intercellular interactions, and intracellular signaling. For instance, MMPs can cleave cell-cell adhesion molecules like E-cadherin, modulate cell-ECM interactions through processing of integrins, convert cytokines precursor such as protransforming growth factor-β and protumor necrosis factor-α, or release cytokines or growth factors like Insulin like growth factor (IGF) from ECM or carrier proteins; moreover cell membrane receptors such as Protease-Activated Receptors (PAR-1) can be processed by MMPs[53,54].

The activity of MMPs can be regulated at different levels:

– Transcription factor binding sites in MMPs gene promoters, Histones Acetylation and DNA methylation regulate MMPs gene expression in response to various stimuli;

– At the post-transcriptional level, RNA-binding proteins and microRNA can regulate the stability of MMPs mRNA;

– Enzymatic activation of the precursor zymogen by cleavage of prodomain;

– Interaction with specific ECM components;

– Inhibition by TIMPs[55].

During tissue injury and repair, the expression levels of many MMPs are regulated by inflammatory cytokines such as TNF-α and IL-1. Beside cytokines, ECM components as well as mechanical stress can modulate MMPs expression.

During cancer, MMPs can be expressed either by tumor cells or stromal and infiltrating inflammatory cells, providing evidence that the adaptative and innate immune system play an important role in the process of tumor progression. Indeed, upregulation of MMPs has traditionally been associated to tumor progression, both at the primary and secondary site, but recent studies reported that some host-derived MMPs have anti-tumorigenic effects[56-58].

Recent studies have also associated MMPs activation to genetic instability. One mechanism proposed to explain the MMPs-dependent genetic instability is associated to nuclear localization of MMPs. Indeed, recent studies observed MMP-2 and MMP-3 in the nuclear compartment and a putative nuclear localization signal was identified in their sequence, as well as in other MMPs[59,60].

These recent findings highlight novel pathways, representing important aspects of MMPs activity and functions in the cells.

I.4.2. Protease-Activated Receptors

Protease-Activated Receptors (PARs) belong to a subfamily of four G proteins-Coupled Receptors (GPCRs) with seven transmembrane domains that acts as sensors of proteases in extracellular environment. PARs are activated by a unique mechanism: proteases activate PARs by proteolytic cleavage within the extracellular N-terminus of their receptors, thereby exposing a novel "cryptic" N-terminal sequence activating the receptor.

Specific residues (about six amino acids) within this tethered ligand domain are believed to interact with extracellular loop 2 and other domains of the cleaved receptor, resulting in intra-molecular activation. This activation process is followed by coupling to G proteins and the triggering of a variety of downstream signal transduction pathways[61] (Fig. 2).

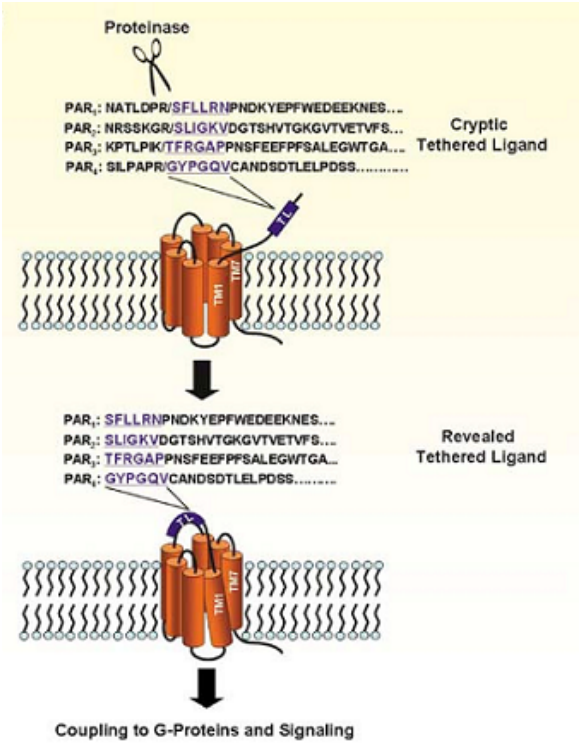


Fig. 2. PARs activation mechanisms by proteases. Partial sequence alignment of PAR1-4 N- terminal sequences, with highlighted the tethered ligand exapeptides.

Stimulation and termination of PAR-mediated signalling is regulated by several mechanisms. The availability of PARs at the cell surface is governed by trafficking of the receptor from intracellular stores, and the signaling properties depend on the presence of G proteins and G protein-coupled receptor kinases (GRKs) that modulate PARs activity[62]. PAR-1, PAR-3, and

PAR-4 are targets for thrombin, trypsin, or cathepsin G. In contrast, PAR-2 is resistant to thrombin, but can be activated by trypsin, mast cell tryptase, factor Xa, acrosin, gingipain, and neuronal serine proteinases.

For PAR-1, PAR-2, and PAR-4, it is well established that short synthetic peptides [PAR-activating peptides (PAR-APs)] designed on their proteolytically revealed tethered ligand sequences can serve as selective receptor agonists and some PAR-APs activate more than one PAR. PAR-3 on its own does not appear to signal and does not respond either to thrombin or to the PAR-AP based on the thrombin-revealed PAR-3 tethered ligand sequence, but this peptide is able to activate either PAR-1 or PAR-2. Further studies have

also provided evidence that a possible interaction with PAR-1 and PAR-4 is necessary for PAR-3 activation.In addition to the cleavage/activation of PARs, proteinases can also negatively regulate PARs function through 'disarming' the receptor by cleavage at a site downstream the receptor-activating site, to remove the tethered ligand. These truncated receptors nonetheless remain responsive to PAR-APs but would be unable to signal in a physiological setting.

In many cases, PARs appear to play a proinflammatory role due to activation of proinflammatory mediators and cytokines. In other instances, a protective and anti-inflammatory role of PARs has been observed[53,63-65].

PAR-1 is the first member of PARs family to be discovered in various cell types (endothelium, platelets, and neutrophils). PAR-1 is coupled with different G proteins and can activate multiple downstream signalling pathways, including the activation of PI3 kinase, Src family tyrosine kinases, the extracellular signal-regulated kinase (ERK)/mitogenactivated protein kinase (MAPK) pathway and signalling to nuclear factor (NF)-kB.

Tumor-expressed PAR-1 play an important role in tumorigenesis of various tissues. Recent

studies show that in breast cancer cell lines, stromal-derived MMP-1 can activate tumor-expressed PAR-1 and promote breast cancer cell migration and invasion[66,67] (Fig.3).

In melanoma and colon cancer cell lines that express MMP-1, an inverse MMP-1/PAR-1 pathway was observed: tumor-derived MMP-1 cleaves microvascular and macrovascular endothelial PAR-1, thus generating a prothrombotic and proinflammatory cell surface[68]. Inhibition of this cross-talk may be a powerful means to prevent tumor-induced endotelial cell activation and thus thrombotic and inflammatory cell adhesion.

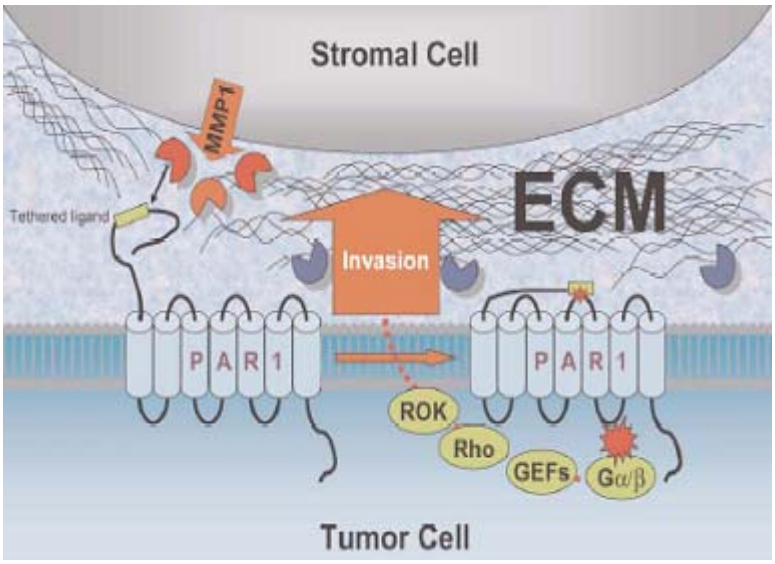


Fig.3. MMP-1-mediated PAR1 activation in breast cancer cells leads to increased cell invasion.

I.4.3. EF-hand Proteins in Signal Transduction Pathways

In all eukaryotic cells, Ca^{2+} ions are important second messengers in a variety of cellular signaling pathways and intracellular Ca^{2+}-binding proteins, containing the specific Ca^{2+} binding motif (helix-loop-helix, called EF-hand, Fig.4), are the key molecules to transduce signaling via enzymatic reactions or modulation of protein/protein interations upon variations in cytosolic Ca^{2+} concentrations.

The EF hand proteins, like calmodulin and S100 proteins, are considered to exert Ca^{2+}-dependent actions in the nucleus, in the cytoplasm and in the extracellular environment.

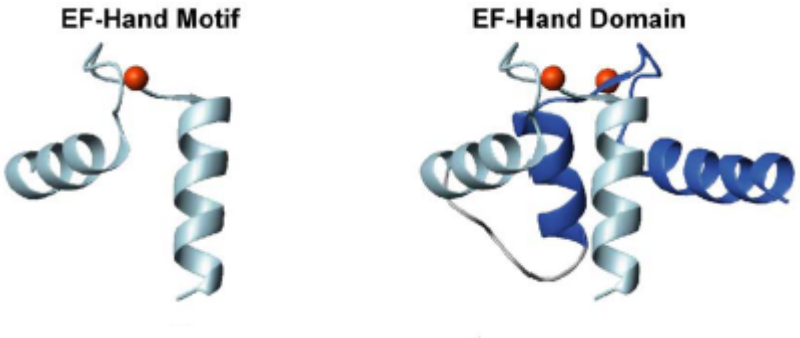


Fig. 4. Structures of Ca^{2+}-loaded EF-hand motif and EF-hand domain, composed by two EF-hand motifs. In red are shown Ca^{2+} ions bound to the loop of each EF-hand motif.

Calmodulin. Calmodulin (CaM) represents the prototypical intracellular Ca^{2+}-sensor containing four Ca^{2+} binding sites in the loops of four canonical EF-hand motifs[69]. It is highly

conserved and widely distributed in all members of the animal and plant kingdoms, fungi, and protozoa, with 100% aminoacid sequence identity among all vertebrates, that synchronize cellular responses to cell activation, resulting from an elevation of $[Ca^{2+}]$[70-73].

In mammals genomes there are three separate genes all coding for a 100% identical CaM molecule of 149 aminoacids, including the N-terminal Met.

CaM is composed of two globular domains, each containing two EF-hand motifs connected by a central helix (Fig. 5).

The pairing of EF-hands enables cooperativity in the binding of Ca^{2+} ions[74]. The two domains share high overall sequence homology (75%), as well as structural similarity in the presence and absence of Ca^{2+} ions. However, two Ca^{2+} ions bind with a tenfold lower affinity (*K*d ~10^{-5} M) to the N-domain than to the C-domain (*K*d ~10^{-6} M). This allows CaM to sense transient Ca^{2+} variations in the cytoplasm over a relatively wide concentration range. Upon Ca^{2+} binding, the linker between the two domains bends round and CaM assumes a more globular shape, ready to wrap around a substrate recognition site. At the same time the two domains undergo conformational changes exposing hydrophobic patches that favour target protein interactions[75] (Fig. 5).

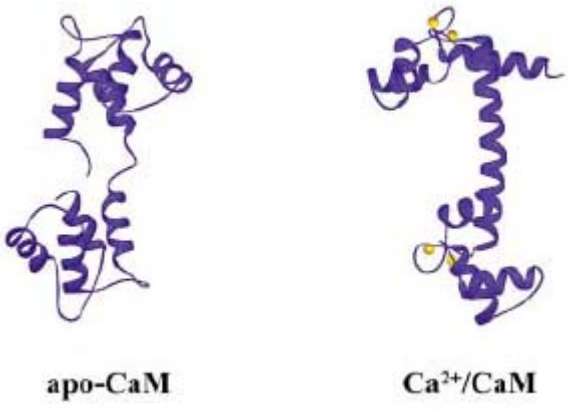


Fig. 5. Structure of CaM in the apo and Ca_4 forms. Upon Ca^{2+} binding, the central helix of CaM, connecting the N- and C-terminal domains, bends round leading to a conformational change in the overall structure of CaM, and to the exposure of hydrophobic residues interacting with target proteins.

CaM can bind to its targets in different ways; in the extended mode interaction its domains interact with different regions of the target. The extended binding mode is also used for targets that bind to apo-CaM. Many of these targets, such as neuromodulin and neurogranin, interact through the IQ motif, which contains the consensus sequence IQxxxRGxxxR[76]. Some IQ motifs bind to CaM in both the absence and presence of Ca^{2+} (e.g., insulin receptor substrate-1, myosin) and in some cases the IQ motif is combined with other CaM binding sequences.

Complexes of CaM with proteins from the family of bHLH transcription factors reveal another CaM binding mode that lead to CaM-induced dimerization of the target.

The flexibility of CaM structure and the different binding modes are the key features that make CaM able to interact with more than hundred different targets, involved in numerous cellular processes such as cell division and differentiation, gene transcription, ion transport by channels, membrane fusion and muscle contraction[77,77,77-79].

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

Table 1. Nomenclature and chromosomal location of S100 family members.

Approved gene symbol	Approved gene name	Previous symbols and aliases	Chromosomal location	Sequence Accession ID
S100A1	S100 calcium binding protein A1	S100A, S100-alpha	1q21	NM_006271
S100A2	S100 calcium binding protein A2	S100L, CaN19	1q21	NM_005978
S100A3	S100 calcium binding protein A3	S100E	1q21	NM_002960
S100A4	S100 calcium binding protein A4	Calvasculin, metastasin, murine placental homolog, calcium placental protein (CAPL), MTSI, p9Ka, 18A2, pEL98, 42A	1q21	NM_002961
S100A5	S100 calcium binding protein A5	S100D	1q21	NM_002962
S100A6	S100 calcium binding protein A6	Calcyclin (CACY), 2A9, PRA, CABP	1q21	NM_014624
S100A7	S100 calcium binding protein A7	Psoriasin 1 (PSOR1), S100A7c	1q21	NM_002963
S100A7A	S100 calcium binding protein A7A	S100A15, S100A7L1	1q21	NM_176823
S100A7L2	S100 calcium binding protein A7-like 2	S100A7b	1q21	–
S100A7P1	S100 calcium binding protein A7 pseudogene 1	S100A7L3, S100A7d	1q21	–
S100A7P2	S100 calcium binding protein A7 pseudogene 2	S100A7L4, S100A7e	1q21	–
S100A8	S100 calcium binding protein A8	Calgranulin A (CAGA), CGLA, P8, MRP8, CFAG, LIAg, 60B8AG	1q21	NM_002964
S100A9	S100 calcium binding protein A9	Calgranulin B (CAGB), CGLB, P14, MRP14, CFAG, LIAg, 60B8AG	1q21	NM_002965
S100A10	S100 calcium binding protein A10	Annexin II ligand (ANX2LG), calpactin I, light polypeptide (CAL1L), p11, CLP11, 42C	1q21	NM_002966
S100A11	S100 calcium binding protein A11	Calgizzarin, S100C	1q21	NM_005620
S100A11P	S100 calcium binding protein A11 pseudogene	S100A14	7q22–q31	–
S100A12	S100 calcium binding protein A12	Calgranulin C (CAGC), CAAF1, CGRP, p6, ENRAGE	1q21	NM_005621
S100A13	S100 calcium binding protein A13		1q21	NM_005979
S100A14	S100 calcium binding protein A14	BCMP84, S100A15	1q21	NM_020672
S100A16	S100 calcium binding protein A16	S100F, DT1P1A7, MGC17528	1q21	NM_080388
S100B	S100 calcium binding protein B	S100-beta	21q22	NM_006272
S100G	S100 calcium binding protein G	Calbindin 3 (CALB3), CaBP9K, CABP1	Xp22	NM_004057
S100P	S100 calcium binding protein P		4p16	NM_005980
S100Z	S100 calcium binding protein Z	S100-zeta	5q13	NM_130772

The first member was identified in 1965 by Moore from bovine brain, and called ‘‘S100’’ because of its solubility in a 100% saturated solution with ammonium sulphate at neutral pH [80,81]. Since then, the expression of S100 proteins has been demonstrated in a diverse spectrum of tissues and involved in the regulation of Ca^{2+} signal transduction pathways.

In vivo and *in vitro* experiments have shown that the S100 proteins can form non covalent homo-and hetero-dimers, with the exception of S100G (Calbindin), which only acts as a Ca^{2+}-buffering protein.

The S100 monomer has two distinct EF-hands, one common to all EF-hand proteins on the C-terminal portion (helixIII-loop-helixIV) and one specific to this family located at the N-terminus (helixI-loop-helixII). Downstream the C-terminal EF-hand region is a stretch of amino acids referred to as the C-terminal extension. Between the two EF-hand domains is the sequence known as the hinge. The C-terminal extension and hinge regions have the most variability between the different proteins and hence are responsible for their specific biological properties.

The two EF-hands in each monomer differ in sequence and mechanisms of calcium coordination. The 12-residue C-terminal EF-hand binds calcium in a similar manner to calmodulin and troponin-C, resulting in a higher calcium affinity site with $K_d \sim 10\text{-}50\ \mu M$.

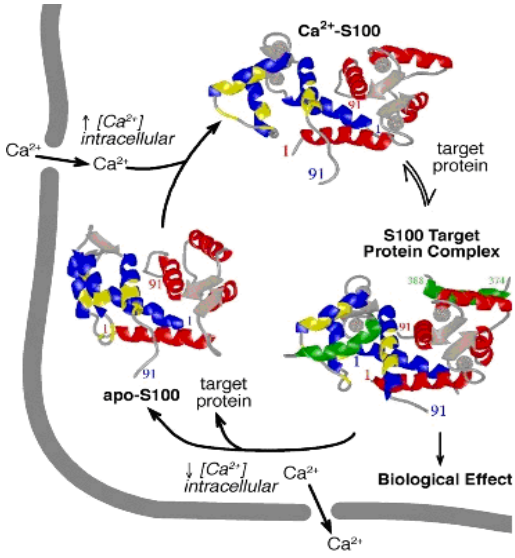


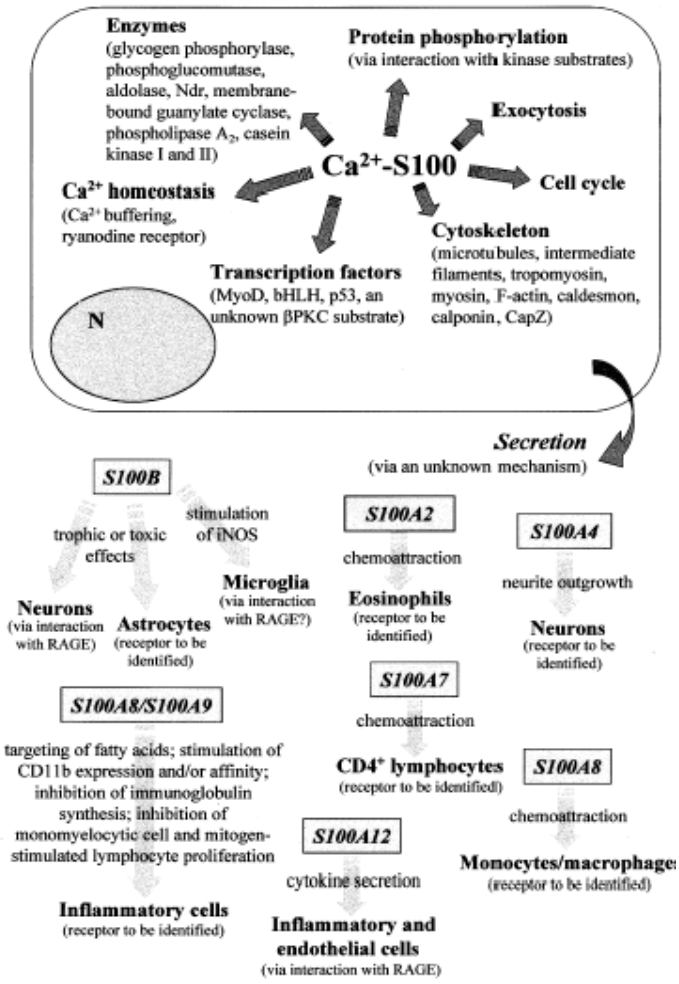
Fig. 6. Ca^{2+}-binding induces structural changes in S100 proteins that allow the exposure of key residues involved in target proteins interaction.

The N-terminal or 'pseudo-canonical' EF-hand is formed by 14 residues and binds calcium with weaker affinity ($K_d \sim 200\text{–}500\ \mu M$). The calcium-induced structural changes in the S100 proteins are responsible of the exposure of residues from helices III and IV in the C-terminal EF-hand, and linker region that facilitate the interaction with target proteins[74,75] (Fig.6). S100A10 is unique within the S100 family, since, upon mutations during evolution, both Ca^{2+}-binding sites are inactive and it is locked in the equivalent of a Ca^{2+}-loaded structure, thus in a permanently activated state[82]. Anyway, Ca^{2+}-indepentent functions were reported for other S100 proteins. The most common binding partners for the apo-S100 proteins are enzymes, and also their abilities to form homo- and heterodimers, as well as some

higher-order complexes is an important Ca^{2+}-independent interaction. For example, S100B forms the tightest dimer (K_d <500 pM) in the calcium-free state[83].

Interest in the S100 proteins comes from their involvement in several human diseases, such as Alzheimer's disease, cancer and rheumatoid arthritis, usually due to modified levels of expression of the S100 members[84,85]. It is well documented that S100 proteins have a broad range of intracellular and extracellular functions (Fig. 7).

Fig. 7. Schematic representation of most of the intra- and extra-cellular activities and pathways regulated by S100 proteins.



Intracellular functions include regulation of different processes: protein phosphorylation (MyoD, neuromodulin, tau protein), enzyme activity and metabolism (NDR kinase, guanylate cyclase, aldolase C), calcium homeostasis (Annexin A6, AHNAK), cytoskeleton dynamics (tubulin, F-actin, intermediate filaments, myosin and tropomyosin), transcription (p53, MyoD), metastasis, cell differentiation and shape, proliferation, and membrane trafficking.

Certain S100 members are released into the extracellular space by an unknown mechanism and can regulate cellular activities in an endocrine, paracrine and autocrine manner[80], by

interacting with RAGE receptor.The Receptor for Advanced Glycation End products (RAGE) has been found to bind some S100 members and transduce signals upon S100 binding[86]. It is a multiligand receptor of the immunoglobulin family, with three domains in the extracellular N-terminal region, a transmembrane region and a cytosolic domain (Fig. 8). So far S100A4, S100B, S100A12, S100A6, S100A11, S100A13, S100A8/A9, S100P has been identified as RAGE ligands.

S100 proteins use different mechanisms for the interaction with RAGE[87]. For example, S100B in the high Ca^{2+} extracellular environment is a homotetramer and upon binding to RAGE, it mediates receptor dimerization. By contrast, S100A12 is found as a hexamer and causes RAGE tetramerization.

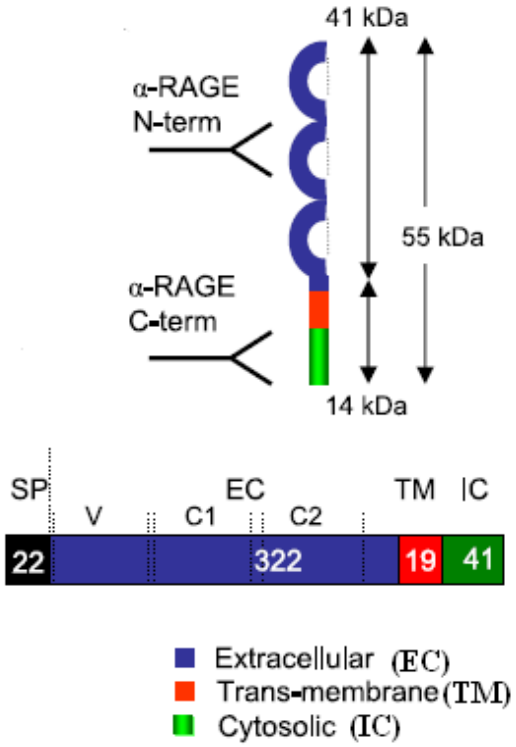


Fig. 8. Domains organization of RAGE receptor. Extracellular S100 proteins can bind different domains of the extracellular region of RAGE.

Recent studies have shown that ligand-activated RAGE can be internalised and targeted to vesicles containing endogenous S100B, which are recycled to the plasma membrane and secreted by a RAGE-dependent mechanism[88].

Some S100s can interact with receptors different from RAGE (still unknown), with non-receptor proteins and extracellular matrix components, regulating RAGE-independent

activities, moreover the oligomerization state of S100 proteins in the extracellular domain showed important functional implications[89].

Extracellular S100s can act as leukocyte chemoattractants, activators for macrophage and other inflammatory cells and modulators of cell proliferation. These functions associate S100 proteins with tissue organization during development and a variety of pathologies such as inflammation, cardiomyopathies, and carcinogenesis.

I.4.4. p53

The intense interest in p53 has generated up to now more than 47300 publications. This interest on p53 comes from its key role in the maintenance of genomic stability and the fact that loss of normal p53 function by mutations occurs in around 50% of human cancers[90]. The p53 is a tumor suppression protein, it induces growth arrest or cell death upon DNA damage or other genotoxic stresses and prevents accumulation of mutations in the genome[91,92]. P53 acts mainly as a transcription factor, regulating the transcription of many genes involved in cellular processes, including cell cycle, DNA repair, apoptosis, angiogenesis, senescence[93].

The major mechanisms involved in modulation of p53 activity are regulation of p53 protein levels, oligomerization, localization and post-translational modifications. P53 structure and domain organization reflects its intricate regulatory mechanisms[94-96].

Human p53 is a 393 amino acids protein which consists of four functional domains: the N-terminal region (1-93), a highly conserved DNA binding domain (residues 94-292), a tetramerization domain (residues 325-356) and finally a regulatory C-terminal region of about 30 residues (Fig. 9).

The N-terminal region of p53 is natively unfolded and consists of an acidic trans-activation domain (TAD) and a proline-rich region[97]. The TAD is a promiscuous binding site for

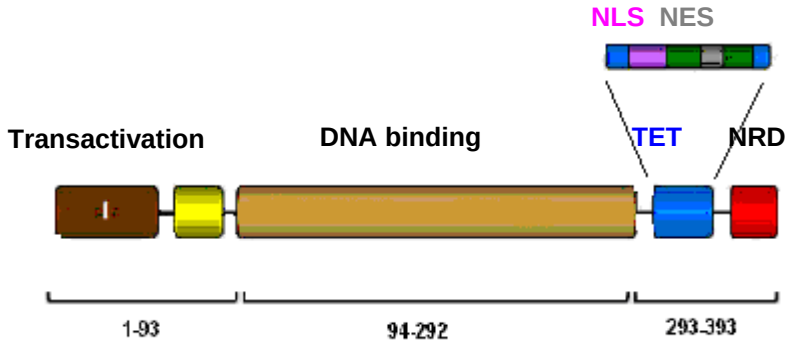


Fig .9. Domains organization of p53 monomer.

different interacting proteins, such as components of the transcription machinery[98-100], the transcriptional coactivators p300/CBP (CREB-binding protein)[101,102], and the negative regulators MDM2/MDM4[103-105], that play key roles in the regulatin of p53 activity. Moreover in the TAD there are sites for posttranslational modifications that further modulate p53 interactions and activity. For example, in response to stress signals several protein kinases phosphorylate multiple N-terminal serine and threonine residues and in this way can modulate the relative affinity for the different proteins that compete for p53 binding[106,107]. The Proline-rich domain contains five PXXP motifs, generally mediating numerous protein-protein interactions through binding to Src homology 3 (SH3) domains, but the exact role of this region is poorly understood.

The DNA binding domain (DBD) is an immunoglobulin-like β-sandwich, subdivided into two structural motifs that bind to the minor groove and major groove of target DNA, respectively[108]. A structural Zinc ion is necessary to maintain thermodynamic stability, DNA binding specificity and avoid protein aggregation[109,110].

P53 bind to specific binding sites in the promoter of its target genes with different affinities, depending on the target gene and on posttranslational modification of p53 DBD, such as Lysine acetylations[111].

The active form of p53 transcription factor is the tetramer, formed through a tetramerizzation domain (TET) in the C-terminal of the protein (Fig. 10). The monomeric TET consists of a short β-strand and an α-helix linked by a turn. Two monomers form a dimer, which is stabilized via an antiparallel intermolecular β-sheet and antiparallel helix packing with central hydrophobic core formed by three key residues (Leu-330, Ile-332, and Phe-341). These dimers associate through their helices to form a four-helix bundle tetramer. The tetramer interface is stabilized largely by hydrophobic interactions, and the key hydrophobic residues are Leu-344 and Leu-348. Dimer are formed cotranslationally on the polysome, whereas tetramers posttranslationally, when p53 concentration increases. The K_d for tetramer formation is ~ 100 nM.

In normally proliferating cells p53 is rapidly degraded by a MDM2-dependent mechanism. MDM2 is an ubiquitin ligase, transcriptionally regulated by p53, mediating ubiquitination of p53 and targeting to the proteasome. The p53-MDM2 interaction can be impaired by phosphorylation of p53 within the MDM2 binding region in response to stress such as DNA damage, mediated by the kinases Chk1, Chk2, ATM and ATR that are activated by genotoxic damages, leading to increased concentration of p53 in the cells.

The C-terminal negative regularoty domain (NRD) is intrinsically disordered but can adopt a

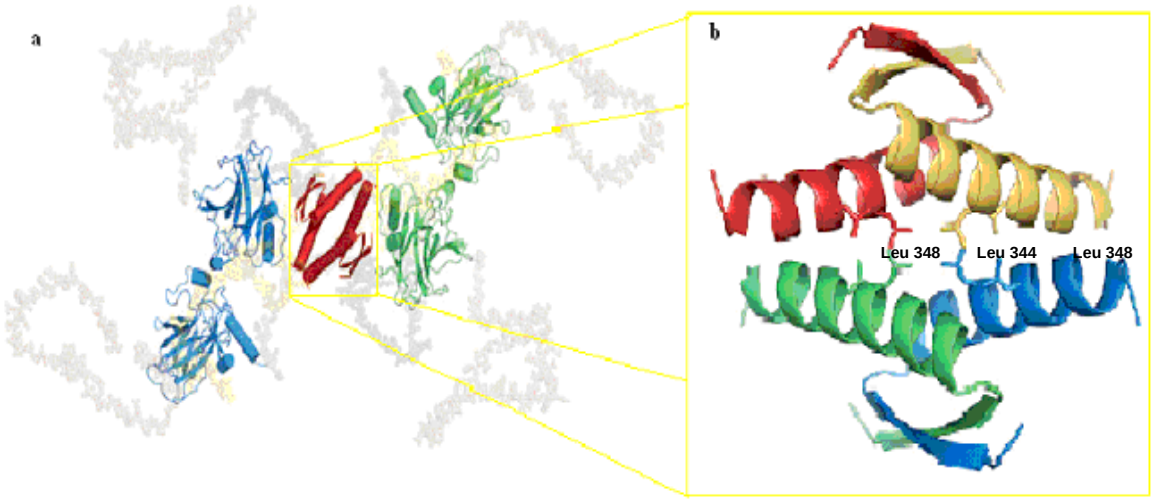


Fig. 10. (a) Model of p53 in solution from small-angle X-ray scattering data [108]. Core and tetramerization domains are shown as cartoon representations, flexible connecting linkers (*gray*), N termini (*pink*), and C termini (*yellow*) are shown as semitransparent space-filled models. (b) Structure of the tetrameric tetramerization domain (PDB id 1C26), composed by a dimer of dimers, stabilized by hydrophobic interactions involving Leu-344 and 348.

helical or a β-turn-like conformation upon binding with regulatory proteins, such as S100 proteins or CREB-binding protein (CBP), respectively.

The NRD is also targeted by posttranslational modification, such as acetylation, ubiquitination, phosphorylation, sumoylation, methylation, and neddylation, that regulate p53 tetramerization, localization, DNA binding, and cellular protein levels.

Since p53 acts mainly as a transcription factor, its localization to the nucleus plays an important role and is strictly regulated by interacting proteins, posttranslational modification and p53 oligomerization[112,113]. P53 contains one bipartite basic nuclear localization signal (NLS) in the linker region between DBD and TET domains and two nuclear export signals (NES), one in the TET domain and one in the N-terminal MDM2 binding region[94].

S100 proteins are important regulators of p53 and act in different way to modulate p53 tetramerization and function, using different binding sites on p53[114].

Mouse models of human cancers have recently demonstrated that continuous expression of a dominantly acting oncogene (H-Ras, K-Ras and Myc) is often required for tumour maintenance[115,116]. Recent studies demonstrated that loss of p53 function may not only play a role in the early stages of tumour development, but also be required for the continued proliferation and survival of an established tumor. These results put new attention on p53 and support efforts to treat human cancers by mechanisms that lead to reactivation of p53.

I.5 AIMS AND TOPICS OF THE RESEARCH

The structural characterization of biomolecular samples, and in particular proteins, require high amounts of purified and native samples, and this step is still the major bottleneck. During my three years PhD, my work has been focused on the expression and characterization of human recombinant proteins, involved in cellular and extracellular pathways.

The physiological function of a protein is strictly related to its three-dimensional structure since in living cells it is regulated by interactions with other proteins and/or macromolecules.

Different techniques can be used to study protein-protein interactions, and different information can be obtained by integrating different approaches.

Solution NMR spectroscopy is a very useful technique, which can provide information about conformational or chemical exchanges, internal mobility and dynamics. In particular, NMR is very efficient to map interaction surfaces of protein/protein complexes.

In the first part of my research, NMR spectroscopy has been integrated with Mass Spectrometry to study the specificity of proteolytic activity of MMP-1 toward the extracellular domain of Protease Activated Receptor-1, which acts as sensor of proteases in the extracellular environment of living cells. Proteolytic cleavage of PAR-1 extracellular domain can lead to receptor activation, or to an irreversibly disarmed receptor. The effect and the signal transduction pathways induced by PAR-1 cleavage depends on the position of the cleavage site. Biochemical studies in cultured breast cancer cells, reported PAR-1 as a MMP-1 target, but the proposed cleavage site does not agree with the known target substrates of MMPs in general. The results of the research have provided insight into the physiologic role of MMP-1 in PAR-1-mediated signal transduction pathways.

In another research work, the specificity of the interaction of S100 proteins, and in particular S100A2 and S100P, with the tumor suppressor protein p53 has been investigated by NMR spectroscopy, native gels and affinity chromatography. This work comes from the observation that different S100 proteins interact with p53 and modulate its tumor suppressor activity in distinct ways, though S100 proteins share high homology in the sequence and in the structure. Since the key role of p53 and S100 proteins in cancer progression, the understanding of their interactions at molecular levels can shine light on some mechanisms involved in p53 activity modulation.

The study of protein-protein interactions has been focused also on the interactions between Calmodulin and its target proteins. Calmodulin is a key protein for the biology of living cells, because it is involved in most signal transduction pathways and is able to interact with a broad

range of target proteins. The structural properties responsible for its versatility has not yet been fully characterised and the knowledge of its structural properties in complex with target proteins and target peptides can lead to some clarifications. Interesting calmodulin targets has been selected by bioinformatic and bibliographic research and one of them has been studied in complex with Calmodulin by NMR spectroscopy techniques.

The last project in which I have been involved is focused on the understanding of the physiological role of proteolytic activity of MMP-13 in early stages of liver fibrosis. There are dissenting opinions on the role of this protease in liver fibrosis. MMP-13 may have different roles in the development and in the recovery of liver fibrosis and these roles can be associated to its ability to cleave different substrates, which can be differently expressed in different stages of the disease. For this purpose, in vivo studies can provide useful information on the effect of MMP-13 on the acute liver inflammation that lead to fibrosis and give some clues on possible MMP-13 targets.

Reference List

1. Puri M, Robin G, Cowieson N, Forwood JK, Listwan P, Hu SH et al. Focusing in on structural genomics: The University of Queensland structural biology pipeline. Biomolecular Engineering 2006; 23(6):281-289.

2. Thornton JM, Todd AE, Milburn D, Borkakoti N, Orengo CA. From structure to function: approaches and limitations. Nat Struct Biol 2000; 7 Suppl:991-994.

3. Yakunin AF, Yee AA, Savchenko A, Edwards AM, Arrowsmith CH. Structural proteomics: a tool for genome annotation. Current Opinion in Chemical Biology 2004; 8(1):42-48.

4. Burley SK. An overview of structural genomics. Nature Structural Biology 2000; 7:932-934.

5. Stevens RC, Wilson IA. Industrializing structural biology. Science 2001; 293(5529):519-520.

6. Todd AE, Marsden RL, Thornton JM, Orengo CA. Progress of structural genomics initiatives: an analysis of solved target structures. J Mol Biol 2005; 348(5):1235-1260.

7. Chandonia JM, Brenner SE. The impact of structural genomics: expectations and outcomes. Science 2006; 311(5759):347-351.

8. Redfern OC, Dessailly B, Orengo CA. Exploring the structure and function paradigm. Curr Opin Struct Biol 2008; 18(3):394-402.

9. Jones DT. Protein structure prediction in the postgenomic era. Curr Opin Struct Biol 2000; 10(3):371-379.

10. Whisstock JC, Lesk AM. Prediction of protein function from protein sequence and structure. Q Rev Biophys 2003; 36(3):307-340.

11. Ganfornina MD, Sanchez D. Generation of evolutionary novelty by functional shift. Bioessays 1999; 21(5):432-439.

12. Marcotte E, Pellegrini M, Thompson M, Yeates T, Eisenberg D. A combined algorithm for genome-wide prediction of protein function. Nature 1999; 402:83-86.

13. Huynen MA, Snel B. Gene and context: Integrative approaches to genome analysis. Advances in Protein Chemistry, Vol 54 2000; 54:345-379.

14. Kolesov G, Mewes HW, Frishman D. SNAPping up functionally related genes based on context information: A colinearity-free approach. Journal of Molecular Biology 2001; 311(4):639-656.

15. Eisenberg D, Marcotte EM, Xenarios I, Yeates TO. Protein function in the post-genomic era. Nature 2000; 405(6788):823-826.

16. Sangar V, Blankenberg DJ, Altman N, Lesk AM. Quantitative sequence-function relationships in proteins based on gene ontology. BMC bioinformatics 2007; 8.

17. Apic G, Gough J, Teichmann SA. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. Journal of Molecular Biology 2001; 310(2):311-325.

18. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. Proceedings of the National Academy of Sciences of the United States of America 1999; 96(8):4285-4288.

19. Ekman D, Bjorklund AK, Frey-Skott J, Elofsson A. Multi-domain proteins in the three kingdoms of life: Orphan domains and other unassigned regions. Journal of Molecular Biology 2005; 348(1):231-243.

20. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS et al. Intrinsically disordered protein. J Mol Graph Model 2001; 19:26-59.

21. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: Implications for structural proteomics. Structure 2003; 11(11):1453-1459.

22. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z et al. The importance of intrinsic disorder for protein phosphorylation. Nucl Acids Res 2004; 32(3):1037-1049.

23. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. Intrinsic disorder and functional proteomics. Biophysical Journal 2007; 92(5):1439-1456.

24. Dunker AK, Brown CJ, Lawson JD, Iakouchova LM, Obradovic. Intrinsic disorder and protein function. Biochemistry 2002; 41:6573-6582.

25. Tompa P. Intrinsically unstructured proteins evolve by repeat expansion. Bioessays 2003; 25(9):847-855.

26. Fuxreiter M, Tompa P, Simon I. Local structural disorder imparts plasticity on linear motifs. Bioinformatics 2007; 23(8):950-956.

27. Weinreb PH, Zhen WG, Poon AW, Conway KA, Lansbury PTJr. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. Biochemistry 1996; 35:13709-13715.

28. Romero P, Obradovic Z, Li XH, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins-Structure Function and Genetics 2001; 42(1):38-48.

29. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. Journal of Molecular Biology 2004; 337(3):635-645.

30. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 1999; 293:321-331.

31. Sugase K, Dyson HJ, Wright PE. Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 2007; 447(7147):1021-1U11.

32. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. Journal of Molecular Biology 2002; 323(3):573-584.

33. Fink AL. Natively unfolded proteins. Curr Opin Struct Biol 2005; 15(1):35-41.

34. Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins-Structure Function and Genetics 2000; 41(3):415-427.

35. Murzin AG. Biochemistry - Metamorphic proteins. Science 2008; 320(5884):1725-1726.

36. Tuinstra RL, Peterson FC, Kutlesa S, Elgin ES, Kron MA, Volkman BF. Interconversion between two unrelated protein folds in the lymphotactin native state. Proceedings of the National Academy of Sciences of the United States of America 2008; 105(13):5057-5062.

37. Robinson CV, Sali A, Baumeister W. The molecular sociology of the cell. Nature 2007; 450(7172):973-982.

38. Nooren IMA, Thornton JM. Diversity of protein-protein interactions. Embo Journal 2003; 22(14):3486-3492.

39. Kim PM, Sboner A, Xia Y, Gerstein M. The role of disorder in interaction networks: a structural analysis. Molecular Systems Biology 2008; 4.

40. Keskin Z, Gursoy A, Ma B, Nussinov R. Principles of protein-protein interactions: What are the preferred ways for proteins to interact? Chemical Reviews 2008; 108(4):1225-1244.

41. Vaynberg J, Qin J. Weak protein-protein interactions as probed by NMR spectroscopy. Trends in Biotechnology 2006; 24(1):22-27.

42. Nooren IMA, Thornton JM. Structural characterisation and functional significance of transient protein-protein interactions. Journal of Molecular Biology 2003; 325(5):991-1018.

43. Levy ED, Erba EB, Robinson CV, Teichmann SA. Assembly reflects evolution of protein complexes. Nature 2008; 453(7199):1262-1U66.

44. Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature 2007; 450(7172):1001-1009.

45. Bourne PE, Addess KJ, Bluhm WF, Chen L, Deshpande N, Feng ZK et al. The distribution and query systems of the RCSB protein data bank. Nucl Acids Res 2004; 32:D223-D225.

46. Aloy P, Russell RB. Structure-based systems biology: a zoom lens for the cell. FEBS Letters 2005; 579(8):1854-1858.

47. Westerhoff HV, Palsson BO. The evolution of molecular biology into systems biology. Nature Biotechnology 2004; 22(10):1249-1252.

48. Kitano H. Systems biology: A brief overview. Science 2002; 295(5560):1662-1664.

49. Kitano H. Computational systems biology. Nature 2002; 420(6912):206-210.

50. Massova I, Kotra LP, Fridman R, Mobashery S. Matrix Metalloproteinases: structures, evolution, and diversification. FASEB J 1998; 12:1075-1095.

51. Verma RP, Hansch C. Matrix metalloproteinases (MMPs): Chemical–biological functions and (Q)SARs. Bioorganic & Medical Chemistry 2007; 15(6):2223-2268.

52. Visse R, Nagase H. Matrix metalloproteinases and tissue inhibitors of metalloproteinases: structure, function, and biochemistry. Circ Res 2003; 92:827-839.

53. Ramachandran R, Hollenberg MD. Proteinases and signalling: pathophysiological and therapeutic implications via PARs and more. British Journal of Pharmacology 2008; 153:S263-S282.

54. Sternlicht MD, Werb Z. How matrix metalloproteinases regulate cell behavior. Annual Review of Cell and Developmental Biology 2001; 17:463-516.

55. Clark IA, Swingler TE, Sampieri CL, Edwards DR. The regulation of matrix metalloproteinases and their inhibitors. International Journal of Biochemistry & Cell Biology 2008; 40(6-7):1362-1378.

56. Orlichenko LS, Radisky DC. Matrix metalloproteinases stimulate epithelial-mesenchymal transition during tumor development. Clinical & Experimental Metastasis 2008; 25(6):593-600.

57. Decock J, Paridaens R, Ye S. Genetic polymorphisms of matrix metalloproteinases in lung, breast and colorectal cancer. Clinical Genetics 2008; 73(3):197-211.

58. Martin MD, Matrisian LM. The other side of MMPs: Protective roles in tumor progression. Cancer Metastasis Rev 2007; 26(3-4):717-724.

59. Kwan JA, Schulze CJ, Wang WJ, Leon H, Sariahmetoglu M, Sung M et al. Matrix metalloproteinase-2 (MMP-2) is present in the nucleus of cardiac myocytes

and is capable of cleaving poly (ADP-ribose) polymerase (PARP) in vitro. FASEB J 2004; 18(2):690-+.

60. Hockenbery DM. MMPs in unusual places. American Journal of Pathology 2006; 169(4):1101-1103.

61. Dery O, Corvera CU, Steinhoff M, Bunnett NW. Proteinase-activated receptors: novel mechanisms of signaling by serine proteases. Am J Physiol 1998; 43(6):C1429-C1452.

62. Darmoul D, Gratio V, Devaud H, Peiretti F, Laburthe M. Activation of proteinase-activated receptor 1 promotes human colon cancer cell proliferation through epidermal growth factor receptor transactivation. Mol Cancer Res 2004; 2(9):514-522.

63. Arora P, Ricks TK, Trejo J. Protease-activated receptor signalling, endocytic sorting and dysregulation in cancer. Journal of Cell Science 2007; 120(6):921-928.

64. Ossovskaya VS, Bunnett NW. Protease-activated receptors: Contribution to physiology and disease. Physiological Reviews 2004; 84(2):579-621.

65. Steinhoff M, Buddenkotte J, Shpacovitch V, Rattenholl A, Moormann C, Vergnolle N et al. Proteinase-activated receptors: Transducers of proteinase-mediated signaling in inflammation and immune response. Endocrine Reviews 2005; 26(1):1-43.

66. Boire A, Covic L, Agarwal A, Jacques S, Sherifl S, Kuliopulos A. PAR1 is a matrix metalloprotease-1 receptor that promotes invasion and tumorigenesis of breast cancer cells. Cell 2005; 120(3):303-313.

67. Kamath L, Meydani A, Foss F, Kuliopulos A. Signaling from protease-activated receptor-1 inhibits migration and invasion of breast cancer cells. Cancer Res 2001; 61(15):5933-5940.

68. Goerge T, Barg A, Schnaeker EM, Poppelmann B, Shpacovitch V, Rattenholl A et al. Tumor-derived matrix metalloproteinase-1 targets endothelial proteinase-activated receptor 1 promoting endothelial cell activation. Cancer Res 2006; 66(15):7766-7774.

69. Chin DH, Means AR. Calmodulin: a prototypical calcium sensor. Trends in Cell Biology 2000; 10(8):322-328.

70. Makalowski W, Zhang JH, Boguski MS. Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences. Genome Research 1996; 6(9):846-857.

71. Friedberg F, Taliaferro L. Calmodulin genes in zebrafish (revisited). Molecular Biology Reports 2005; 32(1):55-60.

72. Friedberg F, Rhoads AR. Sequence homology of the 3 '-untranslated region of calmodulin III in mammals. Molecular Biology Reports 2001; 28(1):27-30.

73. Copley RR, Schultz J, Ponting CP, Bork P. Protein families in multicellular organisms. Curr Opin Struct Biol 1999; 9(3):408-415.

74. Bhattacharya S, Bunick CG, Chazin WJ. Target selectivity in EF-hand calcium binding proteins. Biochim Biophys Acta 2004; 1742:69-79.

75. Grabarek Z. Structural basis for diversity of the EF-hand calcium-binding proteins. Journal of Molecular Biology 2006; 359(3):509-525.

76. Bahler M, Rhoads A. Calmodulin signaling via the IQ motif. FEBS Letters 2002; 513(1):107-113.

77. Jurado LA, Chockalingam PS, Jarrett HW. Apocalmodulin. Physiological Reviews 1999; 79(3):661-682.

78. Valeyev NV, Heslop-Harrison P, Postlethwaite I, Kotov NV, Bates DG. Multiple calcium binding sites make calmodulin multifunctional. Molecular Biosystems 2008; 4(1):66-73.

79. Schaub MC, Heizmann CW. Calcium, troponin, calmodulin, S100 proteins: From myocardial basics to new therapeutic strategies. Biochemical and Biophysical Research Communications 2008; 369(1):247-264.

80. Donato R. S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles. Int J Biochem Cell Biol 2001; 33(7):637-668.

81. Marenholz I, Lovering RC, Heizmann CW. An update of the S100 nomenclature. Biochimica et Biophysica Acta-Molecular Cell Research 2006; 1763(11):1282-1283.

82. Rescher U, Gerke V. S100A10/p11: family, friends and functions. Pflugers Archiv-European Journal of Physiology 2008; 455(4):575-582.

83. Santamaria-Kisiel L, Rintala-Dempsey AC, Shaw GS. Calcium-dependent and -independent interactions of the S100 protein family. Biochemical Journal 2006; 396:201-214.

84. Hatpio R, Einarsson R. S100 proteins as cancer biomarkers with focus on S100B in malignant melanoma. Clinical Biochemistry 2004; 37(7):512-518.

85. Salama I, Malone PS, Mihaimeed F, Jones JL. A review of the S100 proteins in cancer. Ejso 2008; 34(4):357-364.

86. Donato R. RAGE: A single receptor for several ligands and different cellular responses: The case of certain S100 proteins. Current Molecular Medicine 2007; 7(8):711-724.

87. Leclerc E, Fritz G, Weibel M, Heizmann CW, 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 2007; 282(43):31317-31331.

88. Perrone L, Peluso G, Ab Melone M. RAGE recycles at the plasma membrane in S100B secretory vesicles and promotes Schwann cells morphological logical changes. Journal of Cellular Physiology 2008; 217(1):60-71.

89. Leukert N, Vogl T, Strupat K, Reichelt R, Sorg C, Roth J. Calcium-dependent tetramer formation of S100A8 and S100A9 is essential for biological activity. Journal of Molecular Biology 2006; 359(4):961-972.

90. Hollstein M, Sidransky D, Vogelstein B, Harris CC. P53 Mutations in Human Cancers. Science 1991; 253(5015):49-53.

91. Balint E, Vousden KH. Activation and activities of the p53 tumour suppressor protein. British Journal of Cancer 2001; 85(12):1813-1823.

92. Janicke RU, Sohn D, Schulze-Osthoff K. The dark side of a tumor suppressor: anti-apoptotic p53. Cell Death and Differentiation 2008; 15(6):959-976.

93. Balint E, Vousden KH. Activation and activities of the p53 tumour suppressor protein. British Journal of Cancer 2001; 85(12):1813-1823.

94. Joerger AC, Fersht AR. Structural biology of the tumor suppressor p53. Annual Review of Biochemistry 2008; 77:557-582.

95. Bell S, Klein C, Muller L, Hansen S, Buchner J. p53 contains large unstructured regions in its native state. Journal of Molecular Biology 2002; 322(5):917-927.

96. Scoumanne A, Chen X. Protein methylation: a new mechanism of p53 tumor suppressor regulation. Histology and Histopathology 2008; 23(9):1143-1149.

97. Dawson R, Muller L, Dehner A, Klein C, Kessler H, Buchner J. The N-terminal domain of p53 is natively unfolded. Journal of Molecular Biology 2003; 332(5):1131-1141.

98. Di Lello P, Jenkins LMM, Jones TN, Nguyen BD, Hara T, Yamaguchi H et al. Structure of the Tfb1/p53 complex: Insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53. Molecular Cell 2006; 22(6):731-740.

99. Thut CJ, Chen JL, Klemm R, Tjian R. P53 Transcriptional Activation Mediated by Coactivators Taf(Ii)40 and Taf(Ii)60. Science 1995; 267(5194):100-104.

100. Lu H, Levine AJ. Human Taf(Ii)31 Protein Is A Transcriptional Coactivator of the P53 Protein. Proceedings of the National Academy of Sciences of the United States of America 1995; 92(11):5154-5158.

101. Teufel DP, Freund SM, Bycroft M, Fersht AR. Four domains of p300 each bind tightly to a sequence spanning both transactivation subdomains of p53. Proceedings of the National Academy of Sciences of the United States of America 2007; 104(17):7009-7014.

102. Gu W, Shi XL, Roeder RG. Synergistic activation of transcription by CBP and p53. Nature 1997; 387(6635):819-823.

103. Marine JC, Jochemsen AG. Mdmx as an essential regulator of p53 activity. Biochemical and Biophysical Research Communications 2005; 331(3):750-760.

104. Schon O, Friedler A, Bycroft M, Freund SMV, Fersht AR. Molecular mechanism of the interaction between MDM2 and p53. Journal of Molecular Biology 2002; 323(3):491-501.

105. Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ et al. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science 1996; 274(5289):948-953.

106. Bode AM, Dong ZG. Post-translational modification of p53 in tumorigenesis. Nature Reviews Cancer 2004; 4(10):793-805.

107. Toledo F, Wahl GM. Regulating the p53 pathway: in vitro hypotheses, in vivo veritas. Nature Reviews Cancer 2006; 6(12):909-923.

108. Tidow H, Melero R, Mylonas E, Freund SMV, Grossmann JG, Carazo JM et al. Quaternary structures of tumor suppressor p53 and a specific p53-DNA complex. Proceedings of the National Academy of Sciences of the United States of America 2007; 104(30):12324-12329.

109. Duan JX, Nilsson L. Effect of Zn2+ on DNA recognition and stability of the p53 DNA-binding domain. Biochemistry 2006; 45(24):7483-7492.

110. Bullock AN, Henckel J, Fersht AR. Quantitative analysis of residual folding and DNA binding in mutant p53 core domain: definition of mutant states for rescue in cancer therapy. Oncogene 2000; 19(10):1245-1256.

111. Weinberg RL, Veprintsev DB, Fersht AR. Cooperative binding of tetrameric p53 to DNA. Journal of Molecular Biology 2004; 341(5):1145-1159.

112. O'Keefe K, Li HP, Zhang YP. Nucleocytoplasmic shuttling of p53 is essential for MDM2-mediated cytoplasmic degradation but not ubiquitination. Molecular and Cellular Biology 2003; 23(18):6396-6405.

113. Nie LH, Sasaki M, Maki CG. Regulation of p53 nuclear export through sequential changes in conformation and ubiquitination. Journal of Biological Chemistry 2007; 282(19):14616-14625.

114. Fernandez-Fernandez MR, Rutherford TJ, Fersht AR. Members of the S100 family bind p53 in two distinct ways. Protein Sci 2008; 17(10):1663-1670.

115. Grinstein E, Wernet P. Cellular signaling in normal and cancerous stem cells. Cellular Signalling 2007; 19(12):2428-2433.

116. Ventura A, Kirsch DG, McLaughlin ME, Tuveson DA, Grimm J, Lintault L et al. Restoration of p53 function leads to tumour regression in vivo. Nature 2007; 445(7128):661-665.

II METHODOLOGIES IN STRUCTURAL BIOLOGY

Structural genomics (SG) programs have been initiated worldwide with the aim of solving 3D structure of proteins of living organisms, using high throughput (HT) approaches [1]and leading to an exponential increase of the structures deposited in the PDB.

Proteins from eukaryotes remain a difficult class of targets for SG studies, especially large multidomain or membrane-bound proteins or complexes, but they include important biomedical targets, and a large number of them have been subjected to analysis[2]. Recent reports derived from large functional genomics projects seem to indicate that at best only 10-25% of screened proteins can be adapted to an HT approach and can be used for structural characterizations.

These data highlight the importance of protein expression as a key step for the structural biology of interesting proteins not adaptable to HT approaches. For these targets it is important to adopt an interdisciplinary approach and exploit any technique that can help in the screening of the large number of parameters necessary to identify suitable conditions for good samples preparation[3].

Indeed, proteins for structural characterizations are usually required at the milligram level, and quantities in the range of 10-50 mg or higher of pure material need to be produced. One way to increase the rate of success is to express each target protein in different constructs, modified/engineered (for example to eliminate or introduce post-translational modifications, or increase solubility, or attach tags, etc), or labelled (deuterated, ^{15}N, ^{13}C, Sel-Met, etc) forms to facilitate structural characterizations.

II.1. Protein Expression and Purification

II.1.1. Construct Design

The first and crucial step for the expression and characterization of a recombinant protein, is the choice of the construct. In this step, bioinformatic tools are necessary to analise the nucleotidic and aminoacidic sequences and obtain informations useful for the choice of the constructs.

Proteins can have different splicing variants, SNP variants, different isoforms and such informations can be acquired by available genomic and proteins sequences databases and softwares for predictions. The nucleotidic sequences can be downloaded from databases such as GeneBank (http://www.ncbi.nlm.nih.gov/sites/entrez), and Ensembl (http://www.ensembl.org/index.html), informations on the aminoacidic sequence, variants, isoforms, biophysical properties can be found in Swissprot website (http://www.ebi.ac.uk/swissprot/), while informations about predicted or validated SNPs can be searched in databases, such as dbSNPs (http://www.ncbi.nlm.nih.gov/projects/SNP/).

Information acquired in this first step can help in the choice of the cDNA source, indeed many genes are switched off or transcribed at different levels in various cells and tissues and in different conditions (health, diseases).

In order to design different constructs, the target protein properties and domain organization must be known or predicted using different tools that have been developed thanks to the huge amount of data generated in recent years by different genetic, biochemical and structural approaches:

– Transmembrane region can be predicted to design constructs and further strategies for the expression of a soluble, or transmembrane target or a target containing both soluble and transmembrane domains. (http://www.sbc.su.se/~miklos/DAS/, http://www.cbs.dtu.dk/services/TMHMM-2.0/, http://www.ch.embnet.org/software/TMPRED/form.html);

– The presence of signal peptide for the protein localization can be predicted using available tools (http://www.cbs.dtu.dk/services/SignalP/)[4];

– Topological and structural predictions can help in the identification of intrinsically unstructured regions (http://iupred.enzim.hu/, http://bip.weizmann.ac.il/fldbin/findex/), and the prediction of secondary structures (http://npsa-pbil.ibcp.fr/cgibin/npsaautomat.pl?page=/NPSA/npsa_seccons.html) and tertiary structures (http://www.sbg.bio.ic.ac.uk/~3dpssm/);

– Genome browsing is and approach useful to find proteins sharing the same fold and the same consensus sequence within different genomes (www.ncbi.nlm.nih.gov/BLAST), and obtain information useful for best predictions;

– Multiple sequences alignments from different organisms can help in the definition of domain borders, since it is known that domains sequences are more conserved during evolution than linker regions (http://align.genome.jp/);

– Protein domain identification and analysis of protein domain architectures in completely sequenced genomes can be performed using SMART tool (http://smart.embl-

heidelberg.de/)[5];

– Analyse the protein families structures and domain organization help in the definition of a domain (http://www.sanger.ac.uk/Software/Pfam/)

– Known and predicted protein-protein associations, including direct and indirect associations, can be found in the STRING database (http://string.embl.de/);

– N-terminal sequence should respect the "N-end rule", that relates the metabolic stability of a protein to its N-terminal residue[6].

II.1.2. Cloning strategy

Recombinant proteins yield and solubility are highly dependent on the specific protein sequence, as well as on the vector, host cell, and culture conditions used. The more is known about the characteristics of a protein, the more easily it can be expressed, isolated and purified.

For optimal efficiency, various combinations should be simultaneously screened, to determine the conditions that yield the 'best' sample.

The cloning strategy and the expression system are the first steps to be well designed since they will influence the expression protocol. The choice of the expression system depend on many factors, including cell growth characteristics, expression levels, intracellular and extracellular expression, posttranslational modifications, biological activity of the protein of interest[7,8].

For example, to express a protein of prokaryotic origin, the obvious choice is to use E. coli as host, but in case of eukaryotic proteins, different expression systems can be used and the choice will depend on many factors, since each system has its advantages and problems. Currently, many methodological improvement in non-prokaryotic hosts made more accessible and less expensive eukaryotic systems such as yeast, plants, filamentous fungi, insect or mammalian cells grown in culture and transgenic animals[9-11]. Also cell-free protein syntesis has a great potential for the expression of problematic proteins[12], however especially for characterizations that require high amount of labelled samples, such as NMR, the prokaryotic and in particular the *E.coli* expression system is the most widely used. The choice of the expression vector depends on the expression system. For E. coli, a lot of expression plasmids are available for the screening of different expression conditions that can influence the yield of soluble recombinant protein.

Plasmid vectors possess an origin of replication (ori), a gene for antibiotic resistance (usually Amp^R), which allows for selection of cell clones carrying the plasmid, and a multicloning site, for the insertion of the target protein coding sequence.

Classical cloning, using restriction enzymes, typically cannot be adapted to high-throughput approaches, due to the complication of selecting compatible and appropriate restriction enzymes for each cloning procedure and to its multistep process. High-throughput cloning therefore requires procedures which can help the screening of a broad range of conditions in less time, for these reasons new cloning technologies have been developed in recent years (Fig. 1).

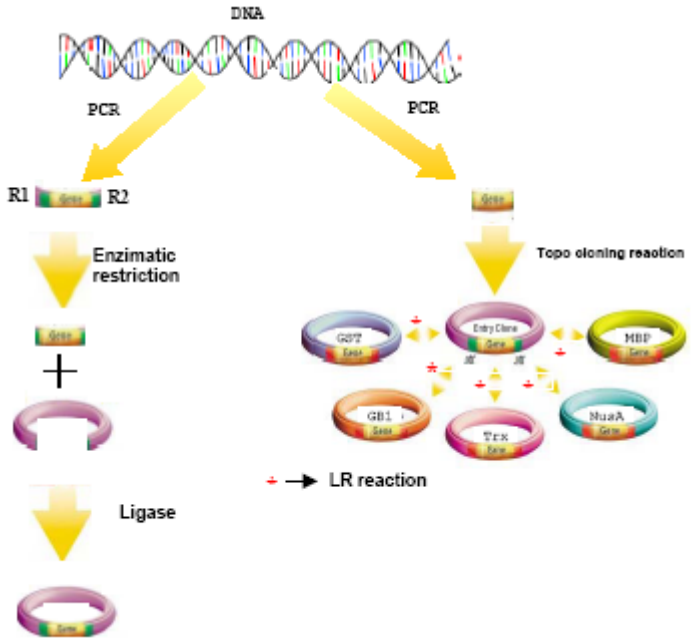


Fig. 1. Schematic representation of ligase-dependent cloning and TOPO cloning (Invitrogen). Classical cloning require multistep procedure, while Gateway technology allow one-step cloning in the pENTR vector, and by site-specific recombination (LR reaction) the target gene can be inserted in different expression vectors.

Landy and co-workers have found a universal cloning method based on the site-specific recombination[13]. Gateway system (Invitrogen) is based on the bacteriophage lambda site-specific recombination system which facilitates the integration of lambda into the *E. coli* chromosome and the switch between the lytic and lysogenic pathways. Gateway system uses this machinery to clone a target gene into different expression vectors, without the time-consuming reactions with restriction enzymes and ligase.

Important elements of an expression plasmid that affect recombinant protein yield and solubility, are: promoter and fusion tag.

Many promoter systems are described as tools for protein expression, especially for E. coli expression system, but only a few of them are commonly used. A useful promoter must be strong, tightly regulated to have a low basal expression level, the induction must be simple and cost-effective, and should be independent on the commonly used ingredients of growth media.

The most used promoter system for the bacterial expression of recombinant proteins is the T7/lac promoter[14]. Genes under the control of T7/lac promoter can be transcribed by T7 RNA polymerase, in presence of lactose. Prokaryotic cells do not produce this type of RNA polymerase, and therefore for the expression of the target proteins can be used only the E. coli strains which has been genetically engineered to incorporate the gene for T7 RNA polymerase, the *lac* promoter and the *lac* operator in their genome. When lactose or a molecule similar to lactose, as Isopropyl β-D-1-thiogalactopyranoside (IPTG), is added to the culture, it displaces the repressor from the *lac* operator. Since there are *lac* operators upstream both the gene encoding the T7 RNA polymerase in the bacterial genome and the target protein in the plasmid, IPTG activates both genes.

T7 RNA polymerase is so selective and active that, when fully induced, almost all of the cell's resources are converted to target gene expression and the desired product can comprise up to 50% of the total cell protein in few hours after induction.

If the basal expression of the recombinant protein must be reduced, as in case of toxic or membrane proteins, or for proteins labelling, host strains containing the pLysS or pLysE vectors can be used. These vectors express the T7 lysozyme, a natural inhibitor of T7 RNA polymerase.

The target protein can be expressed with different fusion partners, which has been developed to increase the expression yield and the solubility of the recombinant proteins, even if sometimes the expression of the native protein could be the best choice[8].

II.1.3. Protein Expression

The screening of different conditions for recombinant protein expression require handling of a huge amount of samples and an high-throughput approach is very useful[1,15,16].

This approach require first the selection of different representative conditions for a preliminary screening, such as fusion tags, promoter systems, E.coli strains, expression temperatures, IPTG concentrations. On the basis of these preliminary results, the expression

protocol can be optimised and, in case of negative results, it is possible to try the expression of mutants, change the cloning strategy, the construct, the expression system. With such an approach it is possible to find good expression conditions for may proteins, anyway some proteins can be difficult to obtain.

The most frequent problems in recombinant protein expression are low expression level, degradation, and insoluble protein expression. To date, there is no generally applicable strategy to solve these problems, but there are different way to increase the rate of success[3,8,17-23].

Besides using different fusion tags and promoter system, a good analysis of the target gene can be useful. Each aminoacid is coded by different codons and the frequency of each codon is different in different organisms. For E. coli expression system, rare codons can be predicted at http://nihserver.mbi.ucla.edu/RACC/. The frequency of these codons reflects the abundance of the tRNAs with the corresponding anticodons, for this reason there are E. coli strains engineered to express extra copies of rare tRNAs that can improve the expression of genes containing rare codons.

Another feature that can increase the yield of expressed protein is the sequence immediately downstream the start codon that can function as a translational enhancer. Some N-terminal tags have been designed to respect these findings.

The stability of the recombinant protein with respect to protease degradation can be improved targeting the recombinant protein in the periplasmic space or in the growth medium, where the concentration of proteases is lower, even if the commercial E. coli strains used for recombinant protein expression are engineered to express lower amount of proteases. The co-expression of the target protein with partner proteins or chaperones can be another way to avoid degradation. Also temperature is a key parameter that influence the protein solubility and degradation kinetic. Indeed, the expression of target gene under the control of cold-shock promoters may have dual effect: decrease the expression of bacterial proteins (and proteases) and increase recombinant protein solubility.

II.1.4. Protein purification

The location of expressed protein within the host will affect the choice of methods for its isolation and purification. For example, a bacterial host may secrete the protein into the growth media, transport it to the periplasmic space, express a cytosolic protein or store it as insoluble in inclusion bodies within the cytoplasm.

For insoluble proteins, the first purification step is the extraction from inclusion bodies. Indeed, the most of the bacterial proteins are removed by different extraction steps with native buffer conditions, while the recombinant protein is extracted from inclusion bodies with a denaturing buffer.

Physical-chemical properties of the recombinant protein also drive the choice of purification protocols, thanks to peculiar properties of the recombinant protein. For instance thermostable protein can be purified by thermal shock. Proteins like S100s can be purified by ammonium sulfate precipitation or, since they expose hydrophobic residues upon Ca^{2+} binding, can be purified by hydrophobic chromatography and eluted by Ca^{2+} removal.

The detection and purification of recombinant proteins can be facilitated by fusion tags that can be used for an affinity purification step (Fig. 2), the most used tags are His tag and GST tag, purified through IMAC chromatography and immobilised Glutatione columns, respectively.

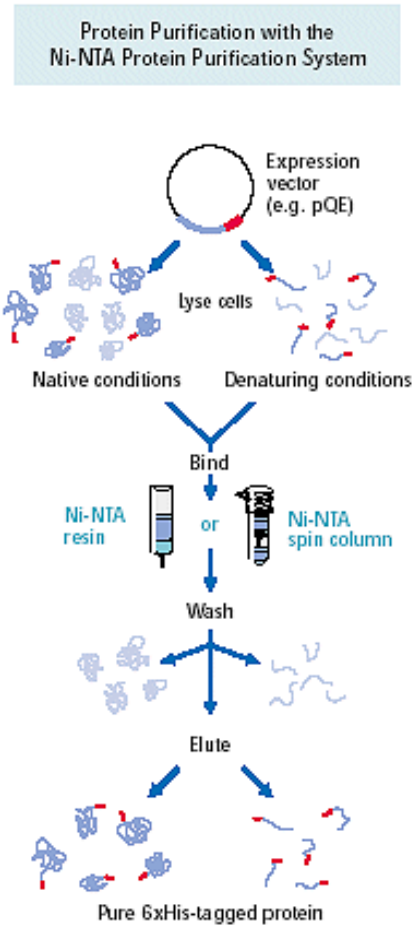


Fig. 2. Schematic representation of His tagged recombinant protein purification. After expression in host cells, the recombinant protein is extracted from lysed cells. Ni-NTA columns can be used ffor purification both in native and denaturing comnditions.

After purification, the fusion tag must be removed from the recombinant protein. Indeed, expression vectors are engineered to express a protease cleavage site between the fusion tag and the recombinant protein.

When the fusion protein is expressed in inclusion bodies, it must be refolded before performing the tag cleavage. Not always this is possible, since fusion tags may interfere with protein folding, therefore other methods must be used to obtain the native protein . If the protein does not contain other methionines in the sequence, one possibility is the CNBr cleavage of starting Methionine.

Structural information of the target protein, as presence of disulfide bridges, or reduced cysteine, or metal binding, can help in the choice of buffer composition, in particular if denaturing agents (DTT, β-Mercaptoethanol, Oxidised/Reduced Glutathione,..), chelating agents (EDTA,EGTA,..) must be used.

II.2 Biophysical Characterizations

II.2.1. Light Scattering

Static and dynamic light scattering represent an approach to studying protein complexes and oligomerization. In static light scattering, the scattering intensity is related to the molecular weight of the protein, in addition to its concentration, the scattering angle, and the wavelength. Dynamic light scattering is based on the auto-correlation of the time-dependent fluctuations of scattered light intensity, which in turn depends upon the diffusion constant. This auto-correlation decays more slowly for slowly diffusing particles and thus, the diffusion constant is extracted from the value of the relaxation time of this function. In the case of ideal spherical particles, this provides a measure of the molecular weight. Light scattering is limited principally by sensitivity, with best results around 1 mg/ml, depending upon the size of the protein or the complex. While sensitivity limits preclude the determination of affinities and association or dissociation rate constants in most cases, light scattering is quite useful in characterizing the stoichiometry of complexes at high concentration. This is very important information for the analysis of data obtained by more sensitive techniques. In fact, it should be pointed out that in many cases, biochemical methods provide the information that protein A interacts with protein B, but the stoichiometry of this interaction is often elusive. Even when crystal structures are available, the stoichiometry of

the complex in the crystal may not correspond to that observed in solution under various conditions.

II.2.2. Mass Spectrometry

Mass spectrometry is an analytical tool used for measuring the molecular mass of a sample.

For large samples such as biomolecules**,** molecular masses can be measured to within an accuracy of 0.01% of the total molecular mass of the sample. Mass Spectrometry can be used for different types of characterization:

– Accurate molecular weight measurements, to determine the purity of a sample, to verify amino acid substitutions, to detect post-translational modifications, to calculate the number of disulphide bridges;

– Reaction monitoring: monitor enzyme reactions, chemical modification, protein digestion;

– Amino acid sequencing: sequence confirmation, characterisation of peptides, identification of proteins by database from proteolytic fragmentation;

– Protein structure: protein folding monitored by H/D exchange, protein-ligand complex formation under physiological conditions, macromolecular structure determination.

Mass spectrometers can be divided into three fundamental parts: the ionisation source , the analyser and the detector.

Sample molecules are ionised in the ionisation source, these ions are extracted into the analyser region of the mass spectrometer where they are separated according to their mass (m) -to-charge (z) ratios (m/z) **.** The separated ions are detected and this signal sent to a data system where the m/z ratios are stored together with their relative abundance for presentation in the format of a m/z spectrum**.**

Many ionisation methods are available, each with its own advantages and disadvantages, and the choice depend on the type of sample under investigation and the mass spectrometer available.

The ionisation methods used for the majority of biochemical analyses are Electrospray Ionisation (ESI) and Matrix Assisted Laser Desorption Ionisation (MALDI).

The main function of the mass analyser is to separate , or resolve , the ions formed in the ionisation source of the mass spectrometer according to their mass-to-charge (m/z) ratios.

There are a number of mass analysers currently available, the better known of which include quadrupoles , time-of-flight (TOF) analysers, magnetic sectors , and both Fourier transform and quadrupole ion traps .

These mass analysers have different features, including the m/z range that can be covered, the mass accuracy, and the achievable resolution.

The detector monitors the ion current, amplifies it and the signal is then transmitted to the data system where it is recorded as mass spectra, where the m/z values of the ions are plotted against their intensities to show the number of components in the sample, the molecular mass of each component, and the relative abundance of the various components in the sample.

MALDI is widely used in biochemical areas for the analysis of proteins, peptides, glycoproteins, oligosaccharides, and oligonucleotides and usually is performed in denaturing conditions.

MALDI is based on the bombardment of sample molecules with a laser light to bring about sample ionisation. The sample is pre-mixed with a highly absorbing matrix compound which transforms the laser energy into excitation energy for the sample, which leads to sputtering of analyte and matrix ions from the surface of the mixture.

The time-of-flight (TOF) analyser separates ions according to their mass(m)-to-charge(z) (m/z) ratios, by measuring the time it takes for ions to travel through a field free region known as the flight tube. The heavier ions are slower than the lighter ones.

In negative ionisation mode, the deprotonated molecular ions ($M-H^-$) are usually the most abundant species, accompanied by some salt adducts and possibly traces of dimeric or doubly charged materials. Negative ionisation can be used for the analysis of oligonucleotides and oligosaccharides. In positive ionisation mode, the protonated molecular ions ($M+H^+$) are usually the dominant species, although they can be accompanied by salt adducts, a trace of the doubly charged molecular ion at approximately half the m/z value, and/or a trace of a dimeric species at approximately twice the m/z value. Positive ionisation is used in general for protein and peptide analyses.

II.2.3. Native gel Electrophoresis

An important tool for the biochemist is the ability to analyze proteins in their native state. Many electrophoresis of proteins and protein:protein complexes in native polyacrylamide gels have been described[24]. This method allows to separate native proteins according to differences in their charge density but not according to molecular weight.

Native gel electrophoresis does not use charged denaturing agents, for this reason some proteins may faile to run in acrylamide native gel electrophoresis.
An alternative technique is the native agarose gel[25], that has the additional advantage of allowing the detection of both positively and negatively charged proteins which migrate toward the anode and cathode, respectively, as well as protein: protein complexes in the same gel. To be sure of the complex formation, the corresponding band can be cut from the gel and the components separated in SDS-PAGE to verify the presence of the two proteins in the complex.

II.2.4. Spectrophotometric Activity assay

Recombinant enzymes, especially those which undergo refolding steps during preparation, have to be checked for activity. Spectrophotometric assays are widely used for determination of enzymatic activity and rely on the difference in molar absorptivity between substrates and products. Since in the UV range the absorbance of proteins may interfere with the measurements, the visible range is preferred for these determinations. Substrates used for these assays are constructed from natural substrates, with new substituents added in order to impart a chromogenic property to the substrate upon the enzymatic reaction, which allows simple quantification.
One example of chromogenic substrate is a commercial chromogenic substrate for spectrophotometric assay of most matrix metalloproteinases. The MMP cleavage site peptide bond is replaced by a thioester bond in this peptide. Hydrolysis of this bond by an MMP produces a sulfhydryl group, which reacts with DTNB [5,5'-dithiobis(2-nitrobenzoic acid), Ellman's Reagent] to form 2-nitro-5-thiobenzoic acid, which can be detected by its absorbance at 412 nm.

II.3. Structural Characterizations

Nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography can provide high-resolution structures of biological molecules such as proteins and nucleic acids and their complexes at atomic resolution. NMR can study molecules in solution, therefore, crystallization is not required, and crystal packing affects may not influence the structure, especially on the surface of a protein. Solution studies should be closer to native-like conditions found in the cell. Since crystals are not needed, protein folding studies can be monitored by NMR spectroscopy upon folding or denaturing of a protein in *real time.* More importantly, denatured states of a biomolecule, folding intermediates and even transition states can be characterized using NMR methods. NMR provides information about conformational or chemical exchanges, internal mobility and dynamics at timescales ranging from picoseconds to seconds, and is very efficient in determining ligand binding, and mapping interaction surfaces of protein/protein, protein/nucleic acid, protein/ligand or nucleic acid/ligand complexes and intramolecular interactions. Improvements in NMR hardware (magnetic field strength, cryoprobes) and NMR methodology, combined with the availability of molecular biology and biochemical methods for preparation and isotope labeling of recombinant proteins, have dramatically increased the use of NMR for the characterization of structure and dynamics of biological molecules in solution. Protein isotope labelling is necessary for NMR analysis because not all atoms are magnetically active.

The nuclei of naturally occurring atomic isotopes that constitute biological molecules have a nuclear spin determined by the spin quantum number (I) and its value depends on the composition of neutrons and protons in each nucleus. Because of the positive charge possessed, the nucleus rotation around its own axis generate a magnetic moment (m). Only atoms with I≠0 can be observed by NMR spectroscopy. For example, ^{12}C is the most abundant isotopes in nature, but it has I=0, for this reason labelling with ^{13}C is performed for biomolecules studied by NMR spectroscopy.

Nuclei which have a spin of one-half, like ^{1}H, or ^{13}C, have two possible spin states: $m = ½$ or $m = -½$ (also referred to as α and β, respectively). The energies of these states are degenerated, hence the *populations* of the two states will be approximately equal at equilibrium.

If a nucleus is placed in a magnetic field, there is interaction between the nuclear magnetic moment and the external magnetic field, then the nuclear spin state aligned with the external magnetic field will be more populated and the different states will have different energies.

When a radiofrequency pulse is applied to match this energy difference, resonance absorption will occur and all nuclei of the same element would resonate at the same frequency. The resonance frequency is affected by the chemical environment of each nucleus, hence nuclei of the same element will have differences in the resonance frequency due to the interference of surrounding electrons, that decrease the magnitude of the effective magnetic field on the nucleus. These differences are called chemical shifts and are higher for more shielded nuclei. The chemical shift value depends also on the applied magnetic field and in order to have chemical shift values normalised on the static magnetic field strength, they are measured in parts per million (ppm).

In a protein, the resonance frequencies of each nucleus vary slightly due to chemical shifts, then a very short radiofrequency pulse is applied which inherently encodes a range of frequencies allowing to induce resonance for the whole frequency spectrum in one experiment (Fourier transform, FT NMR). Transient signals are detected as the system returns to equilibrium. The response obtained from a FT NMR experiment is a superposition of the frequencies of all spins in the molecule as a function of time. In order to obtain the corresponding spectrum as a function of frequency, a Fourier transformation is performed. Fourier transformation is a mathematical operation which translates a function in the time domain into the frequency domain.

Ideally, each distinct nucleus in the molecule experiences a distinct chemical environment and thus has a distinct chemical shift by which it can be recognized. However, in large molecules, such as proteins, the number of resonances can be several thousand and a one-dimensional spectrum inevitably has overlaps. For this reason, proteins NMR spectra cannot be resolved in a conventional one-dimensional spectra (1D) and multi-dimensional nuclear magnetic resonance spectroscopy is required to correlate the frequencies of different nuclei. There are different types of experiments that can detect through-bonds and through-space nucleus-nucleus interactions.

The Heteronuclear Single Quantum Correlation (HSQC) spectrum is a 2D spectrum, where "heteronuclear" refers to nuclei other than ^{1}H. In theory the heteronuclear single quantum correlation has one peak for each H bound to a heteronucleus. Thus the ^{1}H ^{15}N-HSQC spectrum contains the signals of the HN protons in the protein backbone. Since there is only one backbone HN per amino acid, each HSQC signal represents one single amino acid, with the exception of proline, which has no amide-hydrogen due to the cyclic nature of its backbone. Moreover, this HSQC also contains signals from the NH_2 groups of the side chains of Asn and Gln and of the aromatic HN protons of Trp and His.

The acquisition of NMR signals is performed during the relaxation process, which restore the equilibrium of the system. There are two types of relaxation:

– transverse relaxation, due to the interaction between different spins, and measured by the T2 time, which is directly proportional to the molecular weight;

– longitudinal relaxation, due to the interaction between spins and solvent molecules, and is measured by the T1 time.

The main problem in studies of biomolecules with molecular weights above 30 kDa is the fast decay of the NMR signal due to relaxation. Indeed, the line widths in the NMR spectra are inverse proportional to the relaxation rates. Therefore the signal-to-noise in NMR spectra of larger molecules is poor, due to poor resolution and sensitivity. There are NMR methods that can help the acquisition of spectra of large biomolecules, one of them is Transverse Relaxation Optimized Spectroscopy (TROSY). With these improvements high-resolution TROSY-HSQC spectra can be recorded of macromolecules with MWs up to several 100 000 Daltons.

The exchange between two conformations, e.g. free and ligand bound forms of a protein, but also chemical exchange usually gives rise to two distinct NMR signals for a given spin due to different chemical environments in the two exchanging forms. If the exchange rate is slow on the chemical shift time scale, two sets of signals are observed, if the exchange rate is fast on the chemical shift time scale only one signal is observed at an average frequency corresponding to the populations of the two conformations. Intermediate exchange gives rise to very large line width.

Molecular interactions can be very efficiently characterized using very sensitive NMR experiments. Changes in the environment of a spin due to binding of a ligand give rise to chemical shift changes in the NMR spectrum. These changes are usually largest near the binding site. Therefore, the binding surface of a protein with a ligand can be mapped. In addition, from NMR titration experiments dissociation constants can be determined. Due to the relatively high sample concentration, even very weak interactions can be detected. Additional structural information, and long range interactions, as the relative orientation of two protein domains can be measured by the observation of Residual Dipolar Couplings (RDCs).

Dipolar coupled spins are the result of spin/spin interactions through space and depend on the distance between the two spins and the orientation of the internuclear vector with respect to the static magnetic field B_0. The chemical shift difference between ^{1}H-^{15}N (^{1}H-^{15}N J-coupling constant), is different in isotropic or anisotropic conditions. There are different method for

aligning molecules in solution: prepare NMR samples in slightly anisotropic solutions or replace the native metal ions in a molecule with paramagnetic ones, which are able to align the molecules in a magnetic field. In order to measure RDCs, signals of ^{15}N-HSQC experiments must be splitted, and this implies that the numbers of signals doubles with respect to normal HSQC, hence, for proteins this will result in a crowded spectrum. To avoid this problem, two coupled spectra can be acquired, one will contain the component of the coupled magnetization inphase with the external magnetic field and the other the antiphase component. The IPAP (In-Phase AntiPhase) strategy, adopt this technique, moreover it partially overcome the problem of loss of peaks intensity.

Reference List

1. Stevens RC. Design of high-throughput methods of protein production for structural biology. Structure 2000; 8(9):R177-R185.

2. A tour of structural genomics. Nat Rev Genet 2001; 2(10):801-809.

3. Makrides SC. Strategies for achieving high-level expression of genes in Escherichia coli. Microbiological Reviews 1996; 60(3):512-&.

4. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocols 2007; 2(4):953-971.

5. Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P. SMART 5: domains in the context of genomes and networks. Nucl Acids Res 2006; 34:D257-D260.

6. Varshavsky A. The N-end rule: Functions, mysteries, uses. Proceedings of the National Academy of Sciences of the United States of America 1996; 93(22):12142-12149.

7. Yokoyama S. Protein expression systems for structural genomics and proteomics. Curr Opin Chem Biol 2003; 7(1):39-43.

8. Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Applied Microbiology and Biotechnology 2006; 72(2):211-222.

9. Junge F, Schneider B, Reckel S, Schwarz D, Dotsch V, Bernhard F. Large-scale production of functional membrane proteins. Cellular and Molecular Life Sciences 2008; 65(11):1729-1755.

10. Lico C, Chen Q, Santi L. Viral vectors for production of recombinant proteins in plants. Journal of Cellular Physiology 2008; 216(2):366-377.

11. Condreay JP, Kost TA. Baculovirus expression vectors for insect and mammalian cells. Current Drug Targets 2007; 8(10):1126-1131.

12. Klammt C, Schwarz D, Lohr F, Schneider B, Dotsch V, Bernhard F. Cell-free expression as an emerging technique for the large scale production of integral membrane protein. Febs Journal 2006; 273(18):4141-4153.

13. Landy A. Dynamic, Structural, and Regulatory Aspects of Lambda Site-specific Recombination. Ann.Rev.Biochem. 58, 913-949. 1989.

14. Dubendorff JW, Studier FW. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. J Mol Biol 1991; 219(1):45-59.

15. Dieckman L, Gu M, Stols L, Donnelly MI, Collart FR. High throughput methods for gene cloning and expression. Protein Expression and Purification 2002; 25(1):1-7.

16. Folkers GE, van Buuren BN, Kaptein R. Expression screening, protein purification and NMR analysis of human protein domains for structural genomics. J Struct Funct Genomics 2004; 5(1-2):119-131.

17. Scheich C, Kummel D, Soumailakakis D, Heinemann U, Bussow K. Vectors for co-expression of an unrestricted number of proteins. Nucl Acids Res 2007; 35(6).

18. Chatterjee DK, Esposito D. Enhanced soluble protein expression using two new fusion tags. Protein Expression and Purification 2006; 46(1):122-129.

19. Marley J, Lu M, Bracken C. A method for efficient isotopic labeling of recombinant proteins. Journal of Biomolecular NMR 2001; 20(1):71-75.

20. Ohki SY, Kainosho M. Stable isotope labeling methods for protein NMR spectroscopy. Progr NMR Spectrosc 2008; 53(4):208-226.

21. De Marco A, De Marco V. Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning. J Biotechnol 2004; 109(1-2):45-52.

22. Schrodel A, Volz J, De Marco A. Fusion tags and chaperone co-expression modulate both the solubility and the inclusion body features of the recombinant CLIPB14 serine protease. J Biotechnol 2005; 120(1):2-10.

23. Mergulhao FJM, Summers DK, Monteiro GA. Recombinant protein secretion in Escherichia coli. Biotechnology Advances 2005; 23(3):177-202.

24. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. Nature Biotechnology 1999; 17(10):1030-1032.

25. Kim R, Yokota H, Kim SH. Electrophoresis of proteins and protein-protein complexes in a native agarose gel. Analytical Biochemistry 2000; 282(1):147-149.

III. Substrate Specificities of Matrix Metalloproteinase 1 in PAR-1 Exodomain Proteolysis

Nesi A, Fragai M. ChemBioChem, (2007) 8:1367-1369

DOI: 10.1002/cbic.200700055

Substrate Specificities of Matrix Metalloproteinase 1 in PAR-1 Exodomain Proteolysis

Antonella Nesi[a] and Marco Fragai*[a, b]

The signal transduction pathways that are induced by the activation of G protein-coupled proteinase-activated receptors (PARs) play a role in several physiological and pathological processes such as hemostasis, inflammation, angiogenesis, cell adhesion, cancer invasion and metastasis.[1] The molecular basis of PAR-1 triggering has been extensively investigated. In particular, the mechanism by which thrombin, the natural activator of PAR-1, cleaves the protein at the R41–S42 site and unmasks the N-terminal peptide S_{42}FLLRN$_{47}$ has been clarified by NMR spectroscopy, mass spectrometry and molecular biology experiments.[2] Recently, it has been proposed that PAR-1 can also be activated at the same cleavage site by matrix metalloproteinase-1 (MMP-1)[3] and that this non-physiological process can promote invasion and metastasis in several tumor lines where MMP-1 is found to be overexpressed by stromal cells.[4]

However, in our and many other researchers' experience with the recognition of target substrates by MMPs,[5] cleavage at an Arg-Ser peptide bond is unexpected. Even though the broad substrate specificity of MMPs makes it difficult to safely predict the cleavage sites, an amino acid residue with a lipophilic side chain that is downstream of this cleavage site usually fits much better into the S1′ pocket of the enzyme, and a Ser residue is quite unfit for this interaction. In this work, we sought to prove—or disprove—the cleavage of PAR-1 at the R41–S42 bond by MMP-1. These findings might open up new prospects in the understanding of the biology and pharmacology of this class of receptors.

The PAR-1 exodomain A26–L103 is recognized and activated by its physiological partner thrombin.[2] The construct A26–L103 is therefore a biologically meaningful model to study this interaction with MMP-1, and to verify new hypotheses on the activation mechanism. The degradation of the N-terminal domain of PAR-1 by thrombin, which occurs at submillimolar concentrations, has been already monitored by NMR spectroscopy,[2] and was thus used as a reference in the present work.

An analysis of the thrombin- and MMP-1-mediated proteolysis of PAR-1 was carried out in parallel by using 40 μM samples of the construct A26–L103 in a buffer that contained 10 mM Tris (pH 7.2), 5 mM $CaCl_2$, 0.1 mM $ZnCl_2$ and 0.3 M NaCl (Figures S1 and S2 in the Supporting Information). The NMR spectra that were acquired to monitor the thrombin–PAR-1 interaction nicely reproduced the already published data, where the proteolysis at the scissile bond R41–S42 and the structure of the cleaved peptides were well characterized (Figure 1 A). On the contrary, the ^{1}H,^{15}N HSQC spectra of the ^{15}N-enriched PAR-1 exodomain (collected at different times) after the addition of recombinant MMP-1, showed a completely different pattern of signal changes (Figure 1 B). In addition, the proteolysis that is catalyzed by MMP-1 is much slower than that of thrombin, where a few minutes in presence of 0.2 μM of enzyme at 278 K were enough to process all of the PAR-1 polypeptide that was present in solution. The PAR-1 proteolysis by MMP-1 was monitored at 298 K for 53 h (Figure S2). The higher enzyme concentration (up to 3 μM) and the higher temperature were needed to accelerate the proteolytic cleavage of PAR-1 to a suitable extent.

The profile of the enzymatic digestion was determined by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (MS) on aliquots of the original ^{15}N-labelled samples with a degree of enrichment of 97.1 %. The analysis of the PAR-1 exodomain A26–L103 and of the resultant proteolytic products, which was performed by MS confirmed important differences between the proteolytic cleavage of PAR-1 by thrombin and by MMP-1. Most of the peptides that were de-

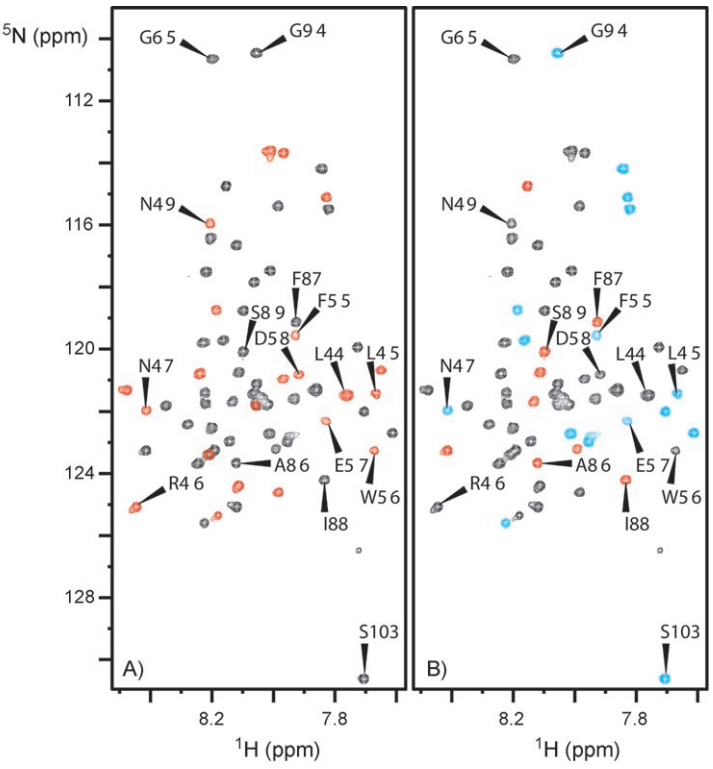


Figure 1. A) ^{1}H,^{15}N HSQC spectra at 278 K of the native A26–L103 PAR-1 exodomain (40 μM); the resonances that are affected by the proteolytic activity of thrombin are in red. B) ^{1}H,^{15}N HSQC spectra at 278 K of the native A26–L103 PAR-1 exodomain (40 μM); the resonances that are affected by the faster full-length active MMP-1 hydrolysis are in red, and the resonances that are affected by the slower hydrolysis are in cyan.

[a] *Dr. A. Nesi, Dr. M. Fragai*
Magnetic Resonance Center, University of Florence
Via L. Sacconi 6, 50019 Sesto Fiorentino, Florence (Italy)
Fax: (+39) 0554574253
E-mail: fragai@cerm.unifi.it

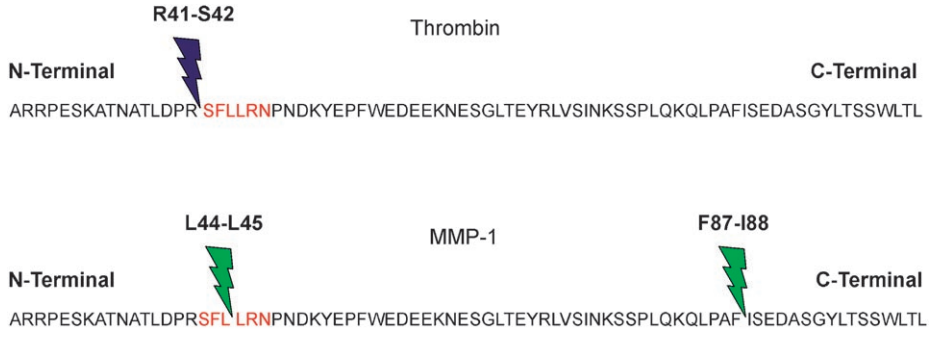
[b] *Dr. M. Fragai*
Department of Agricultural Biotechnology, University of Florence
Via Maragliano 75–77, 50144 Florence (Italy)

Supporting information for this article is available on the WWW under http://www.chembiochem.org or from the author.

Table 1. MALDI/MS analysis of PAR-1 exodomain and of its cleavage products after digestion with thrombin and MMP-1 respectively.

PAR-1(A26–L103)			PAR-1(A26–L103) + thrombin			PAR-1(A26–L103) + MMP-1		
$[M-H]^+$	Fragment analysis		$[M-H]^+$	Fragment analysis		$[M-H]^+$	Fragment analysis	
9053.84	A26–L103	8948.9 + (107-N) 97.1 %	9053.84	A26–L103	8948.9 + (107-N) 97.1 %	9053.84	A26–L103	8948.9 + (107-N) 97.1 %
5438.32	A26–R70	5371.8 + (68-N)	7262.98	S42–L103	7183.9 + (80-N)	7311.27	A26–F87	7224 + (90-N)
3631.95	L71–L103	3595.1 + (39-N)	1809.85	A26–R41	1782.9 + (27-N)	6911.70	F45–L103	6836.5 + (77-N)
			5437.38	A26–R70	5371.8 + (68-N)	5168.02	L45–F87	5111.6 + (60-N)
			3631.90	L71–L103	3595.1 + (39-N)	2159.03	A26–L44	2130.3 + (30-N)
						5436.56	A26–R70	5371.8 + (68-N)
						3631.97	L71–L103	3595.1 + (39-N)
						3295.49	F45–R70	3259.4 + (38-N)
						1891.90	L71–F87	1870.2 + (22-N)

tected in the spectra appeared as single-charged monoprotonated molecular ions at *m/z* $[M+H]^+$. The fragment analysis is reported in Table 1. The progress of PAR-1 digestion by thrombin was followed by periodically removing aliquots after the addition of the enzyme. An incubation time of a few minutes with 0.2 μM of thrombin was enough to completely cleave the native polypeptide at a single cleavage site between R41 and S42 (Scheme 1). After 5 min, the peak of the original peptide A26–L103 disappeared, and only two peaks of *m/z* 1810 and *m/z* 7263, which correspond to the N-terminal A26–R41 and the C-terminal S42–L103 fragment, respectively, remained in the spectra. These two fragments were not further degraded, even after long incubation times. The fast enzymatic degradation of PAR-1 by thrombin under these experimental conditions nicely match with the reported data.[2,7] As was also demonstrated by NMR spectroscopy, the hydrolytic activity of MMP-1 toward PAR-1 was much slower than that of thrombin.



Scheme 1. Cleavage sites of thrombin and of MMP-1 on the A26–L103 PAR-1 exodomain. The reported functional hexapeptide SFLLRN is highlighted in red.

For the MMP-1-treated samples, a mass peak of *m/z* 7311, which is related to a cleavage between F87 and I88 and two other peaks of *m/z* 2159 and *m/z* 6912, which correspond to the N-terminal A26–L44 and to the C-terminal L45–L103 fragment, respectively, were detected in the spectra after an incubation time of 15 min at 278 K. The existence of these two independent cleavage sites was also confirmed by the presence of mass peaks that correspond to degradation products from the already-formed fragments L45–L103 and A26-F87, as well as the peptide A26–R70, which was present as a contaminant also in the absence of the enzyme. Even after long incubation times, peaks that corresponded to the thrombin cleavage sites could not be found in the spectra. The same results were obtained by using the catalytic domain of MMP-1 instead of the full-length active protein. From a careful analysis of the $^1H,^{15}N$ HSQC, two kinetically distinct hydrolytic processes could be detected; a first set of signals disappeared faster and a second set more slowly, this was accompanied by the reappearance of new signals, which also had different time courses.

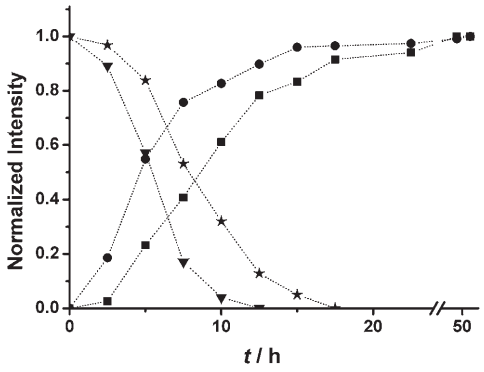


Figure 2. Normalized intensity profiles of 2D $^1H,^{15}N$ HSQC signals for I88 (▼), L45 (★) and for two new signals of the proteolytic fragments (●, ■). Two kinetically distinct hydrolytic processes are detected.

The intensity changes of selected signals that belong to the two sets are shown in Figure 2.

Stepwise proteolysis at the peptide bonds F87–I88 and L44–L45 is consistent with the $^1H,^{15}N$ HSQC spectra, according to the reported assignment. A correlation of the cleavage sites with the two kinetically distinct processes could also be established, because the faster hydrolysis affected the resonances that correspond to F87 and I88, while the loss of the L45 signal was associated with the slower one. The other degradation products that were identified by MS could not be detected by NMR due to their low concentration, even up to 53 h after the addition of the enzyme. The analysis of the intensity profiles, which was performed on isolated NMR signals provided half-lives of 4.9 h and 8.6 h for the two processes, respec-

tively (Figure S3). Past experience with MMPs also suggests that the cleavage at the F87–I88 and L44–L45 sites is much more plausible, because it is widely reported that a hydrophobic amino acid (L45 or I88 in this case) is usually present at the P_1' position to interact with the S_1' binding site of the MMP-1 enzyme.[4] This specificity is also in agreement with the "snapshots" that were recently obtained along the catalytic mechanism of matrix metalloproteinases.[6]

The observation of a different specificity between thrombin and MMP-1 for PAR-1 raises several questions about the mechanism by which MMP-1 activates the receptor. As previously reported for other proteases, proteolysis downstream of what is widely believed to be the functional SFLLRN hexapeptide, which is the tethered ligand of the PAR-1 exodomain would be expected to permanently inactivate the receptor.[7] Consequently, the cleavage at the F87–I88 position should be followed by PAR-1 desensitization. The biological data on the capability of MMP-1 to activate the receptor[3] suggest that this cleavage site, which is near the cellular membrane, might not be easily accessible in vivo for the enzymatic hydrolysis by MMP-1.

The cleavage that occurs at L44–L45, right in the middle of the SFLLRN hexapeptide is much more interesting. The loss of the first three amino acids would be expected to strongly decrease or even to abolish the biological activity because several structure–activity studies that were performed by using libraries of peptides and receptor mutants have established that the amino acids at position 2, 4, and 5 of the hexapeptide are most important for PAR-1 activation.[1,8] Apparently, this is not the case.

An interesting observation is that the biological activity of synthetic hexapeptide analogues, though decreased, is still present in shorter fragments.[8] This weak activity is probably related to a reduced affinity for the receptor that is caused by the loss of intermolecular interactions from the missing amino acids. Unlike the free peptides that were used in structure–activity studies, in the proteolytic unmasking of the N-terminal SFLLRN ligand, the tethering contribution to the binding that is due to the link between the residual PAR-1 exodomain and the receptor has to be taken into account. This phenomenon, which has been widely exploited in drug discovery to increase the affinity of weak ligands,[9] could explain an at least partial PAR-1 activation by the MMP-1-mediated cleavage at the L44–L45 site.

Although the low efficiency in PAR-1 proteolysis with respect to thrombin could call into question the role of MMP-1 in the invasiveness of breast cancer cells, the involvement of interstitial collagenase in this pathological process has been demonstrated.[3] Currently, the reported bimodal activity of thrombin on tumor cell migration (promoting at low, inhibiting at high concentration)[3] suggests that the low cleavage efficiency of MMP-1 might be pathologically relevant.

Experimental Section

Expression and purification protocols of PAR-1 (A26–L103), of the full-length active human fibroblast collagenase G99–N463, and of its catalytic domain N106–G261 are described in detail in the Supporting Information. The enzymatic activity of MMP-1 (400 U μg^{-1}) was determined by using a colorimetric assay (Biomol cat. P-125). Human thrombin was purchased from Sigma–Aldrich. Protein concentrations were determined by a Bradford Assay (Pierce). For all NMR and MS experiments, the PAR-1 exodomain (A26–L103) was resuspended in Tris (10 mM, pH 7.2), NaCl (0.3 M), $CaCl_2$ (5 mM), $ZnCl_2$ (0.1 mM) at a final concentration of 40 μM. A PAR-1 concentration of 40 μM was used in order to ensure the stability of the samples for the extended times that were required to monitor the proteolytic activity of the enzymes, and to avoid the precipitation that easily occurs at higher concentrations.

The mass spectrometry investigation was performed by incubating 40 μM PAR-1 (A26–L103) solutions with thrombin (0.2 μM) and with either full-length active MMP-1 (0.5 μM) or its catalytic domain at 278 K in a buffer that contained Tris (10 mM pH 7.2), $CaCl_2$ (5 mM), $ZnCl_2$ (0.1 mM) and NaCl (0.3 M). Samples at different reaction times were collected and the reaction was stopped by acidification with 0.25% TFA. The reaction samples were purified by Zip Tip (Eppendorf) and analyzed on a Bruker Ultraflex TOF/TOF.

The NMR analysis of the thrombin-mediated proteolysis was carried out by recording the $^1H,^{15}N$ HSQC spectra at 278 K. The proteolytic activity of MMP-1 was investigated by adding increasing concentrations of the enzyme (up to 3 μM) to a PAR-1 sample and by monitoring the reaction at 298 K.

All the spectra were acquired on a Bruker DRX 800 MHz spectrometer, equipped with a TXI cryo-probe.

Acknowledgements

This work was supported by Ente Cassa di Risparmio di Firenze; Fondazione Monte dei Paschi di Siena 2004; PRIN 2005 (MIUR), Prot. N. 2005039878, EU projects: LSHG-CT-2004–512077; "ORTHO AND PARA WATER" no. 00503; LSHB-CT-2005-019102; FIORGEN.

Keywords: enzymes · mass spectroscopy · metalloproteases · NMR spectroscopy · receptors

[1] S. R. Macfarlane, M. J. Seatter, T. Kanke, G. D. Hunter, R. Plevin, *Pharmacol. Rev.* **2001**, *53*, 245 – 282.
[2] S. Seeley, L. Covic, S. L. Jacques, J. Sudmeier, J. D. Baleja, A. Kuliopulos, *Chem. Biol.* **2003**, *10*, 1033 – 1041.
[3] A. Boire, L. Covic, A. Agarwal, S. Jacques, S. Sherifl, A. Kuliopulos, *Cell* **2005**, *120*, 303 – 313.
[4] D. Pei, *Cancer Cell* **2005**, *7*, 207 – 208.
[5] B. E. Turk, L. L. Huang, E. T. Piro, L. C. Cantley, *Nat. Biotechnol.* **2001**, *19*, 661 – 667.
[6] I. Bertini, V. Calderone, M. Fragai, C. Luchinat, M. Maletta, K. J. Yeo, *Angew. Chem.* **2006**, *118*, 8120 – 8123; *Angew. Chem. Int. Ed.* **2006**, *45*, 7952 – 7955.
[7] D. Loew, C. Perrault, M. Morales, S. Moog, C. Ravanat, S. Schuhler, R. Arcone, C. Pietropaolo, J. P. Cazenave, A. Van Dorsselaer, L. O. Lanza, *Biochemistry* **2000**, *39*, 10812 – 10822.
[8] R. M. Scarborough, M. A. Naughton, W. Teng, D. T. Hung, J. Rose, T. K. H. Vu, V. I. Wheaton, C. W. Turck, S. R. Coughlin, *J. Biol. Chem.* **1992**, *267*, 13 146 – 13 149.
[9] S. L. Johnson, M. Pellecchia, *Curr. Top. Med. Chem.* **2006**, *6*, 317 – 329.

Received: February 4, 2007
Published online on June 28, 2007

CHEMBIOCHEM

Supporting Information

for

Substrate Specificities of Matrix Metalloproteinase 1 in PAR-1 Exodomain Proteolysis

Antonella Nesi, Marco Fragai*

Protein Expression and Purification: The N-terminal extracellular domain fragment of PAR-1 (A26-L103) was amplified by PCR from the full-length active cDNA clone (RZPD) by using primers required by Gateway TOPO Cloning Reaction (Gateway Cloning Technology, Invitrogen): Forward: CACC ATG GGG CCG CGG CGG CTG CTG - Reverse: CTA GAG TGT CAG CCA GGA GCT. The purified PCR product was cloned into the pENTR/TEV/D-TOPO vector to obtain the pENTR clone and sequenced. Then the PAR-1 construct was inserted into the pDEST 17A vector by an LR clonase reaction to create the expression vector containing the His tag – TEV cleavage site – PAR-1(A26-L103) construct. The recombinant PAR-1 protein was expressed in *E. coli* strain Gold. The cells were grown at 37°C until OD_{600} of 0.8 was reached. The protein expression was induced by addition of 1 mM IPTG. The cells were allowed to growth further at 30°C for 18-20 h and then harvested by centrifugation. The inclusion bodies, containing the recombinant protein, were solubilized in 20 mM Tris, 0.5 M NaCl, 8 M Urea, 5 mM imidazole (pH 7.5) and purified with Ni^{2+}-loaded HiTrap Chelating column (Amersham). A yield of 50 mg of ^{15}N labelled purified protein per liter of culture was obtained. The recombinant protein was cleaved at starting methionine to remove the N-terminal tag using CNBr/70% Formic Acid, purified by RP-HPLC as described,[1] lyophilised and stored at –80°C.

The full-length active human fibroblast collagenase G99-N463 and the catalytic domain N106-G261 were expressed in E. coli. The cDNA was cloned into the pET21 vector (Novagen) using *Nde I* and *Xho I* as restriction enzymes. The *E. coli* strain BL21 Gold cells, transfected with the above vector, were grown in 2 × YT media at 37°C. The protein expression was induced during the exponential growth phase with 0.5 mM of IPTG. Cells were harvested for 4 h after induction. The cells were lysed by sonication and the inclusion bodies, containing the MMP-1, were solubilized in 6 M guanidine hydrochloride and 10 mM DTT. Then the solution was diluted, without purification, in a buffer containing 50 mM Tris (pH 7.2), 10 mM $CaCl_2$, 0.1 mM $ZnCl_2$, 0.3 M NaCl, 2.3 M guanidine hydrochloride, 20% glycerol, 2.5 mM GSSG, and 2.5 mM GSH at pH 7.8. The refolded protein was exchanged, by dialysis, against a buffer with 50 mM Tris (pH 7.2), 10 mM $CaCl_2$, 0.1 mM $ZnCl_2$, 0.3 M NaCl. The protein was purified on the Hitrap Q column (Pharmacia).

---

[1] A. Kuliopulos, C. T. Walsh, *J.Am.Chem.Soc.* **1994,** *116,* 4599-4600.

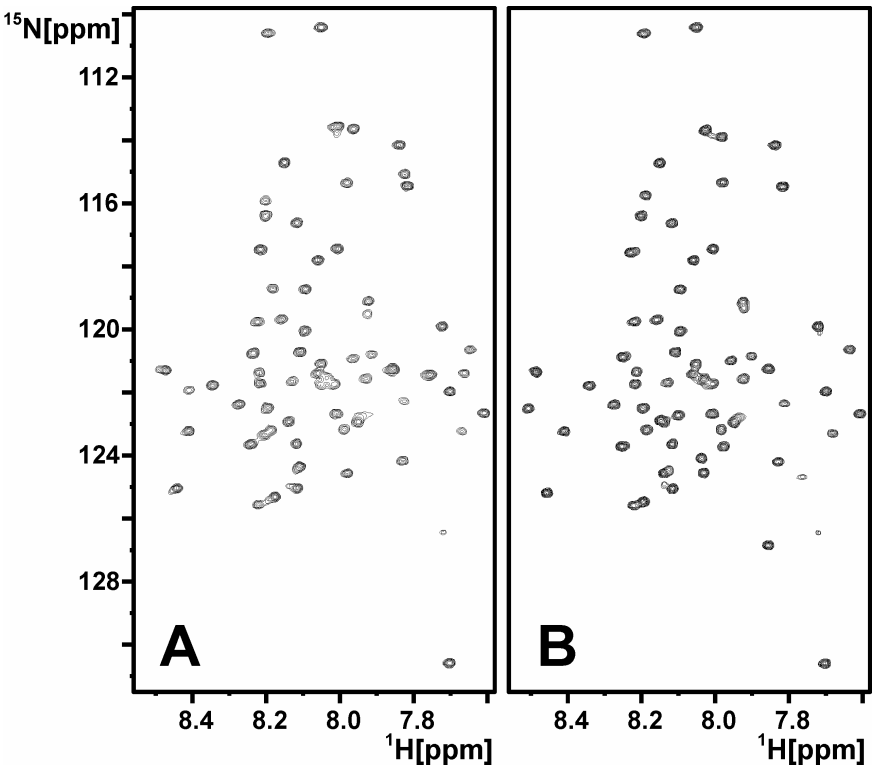


Figure S1. ^{1}H-^{15}N HSQC spectra at 278 K of the native A26-L103 PAR-1 exodomain (40 μM) before (A) and 24 hours after the addition of thrombin 0.2 μM (B).

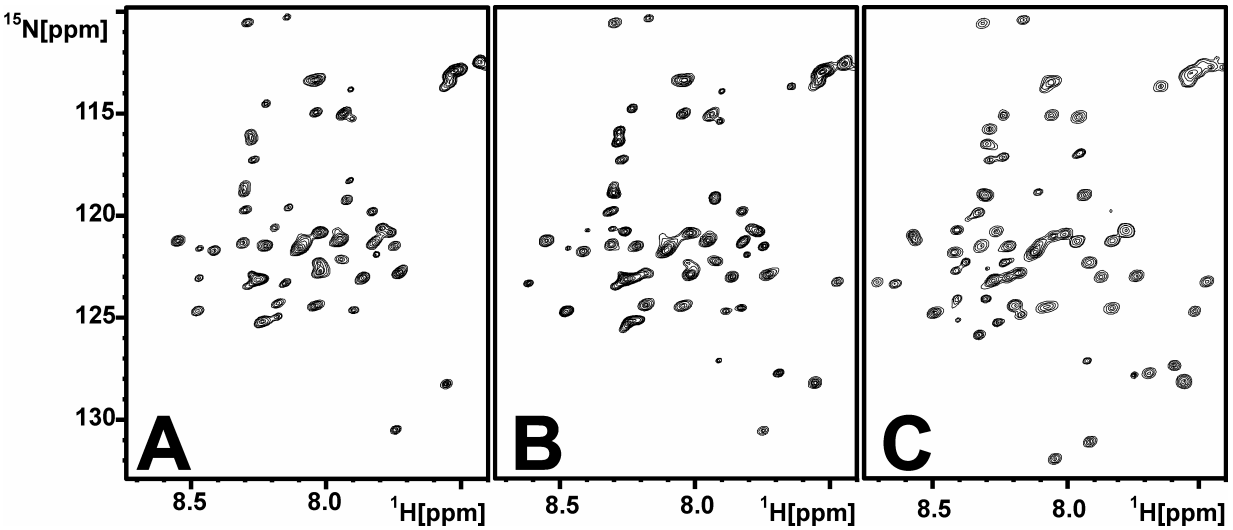


Figure S2. ^{1}H-^{15}N HSQC spectra at 298 K of the native A26-L103 PAR-1 exodomain (40 μM) before (A), 7.5 h after (B) and 53 h after (C) the addition of active full length MMP-1 (3 μM).

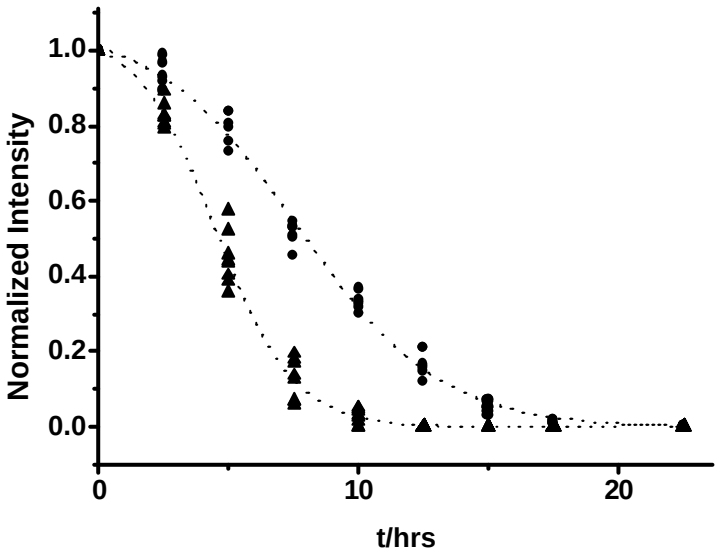


Figure S3. Normalized intensity profiles of the 2D ^{1}H,^{15}N HSQC signals related to the two kinetically distinct hydrolytic processes with half-lives of 4.9 h (7) and 8.6 h (,) respectively.

IV. Proteolytic anti-inflammatory activity of catalytic domain of human MMP-13 on murine models of liver fibrosis

Abstract

AIM: On the basis of the already known double roles of MMP13 in the promotion of early stages of inflammation induced by liver injury and in the fibrosis resolution in rodent models of liver fibrosis, the effects of treatment of murine models of CCl_4-induced liver fibrosis with catalytic domain of recombinant human MMP13 were evaluated. METHODS: Mice and rats were randomly divided in four groups: control, treated with CCl_4, treated with MMP13, and treated with both CCl_4 and MMP13. Liver tissue sections from each group were stained with H&E for morphological examination and with Sirius Red for collagen visualization. RESULTS: The non-collagenolytic proteolytic activity of cat-MMP13 plays a key role in the reduction of the inflammatory process induced by CCl_4 treatment, by reducing the amount of necrotic cells and inflammatory cells in the hepatic tissue.

Key words: Liver fibrosis – CCl_4 – MMP13 – Inflammation – Necrosis.

Introduction

During liver fibrosis there is a pathologic imbalance between the formation and the degradation of extracellular matrix. The main actors of this remodeling are the stellate cells (HSC) that produce collagen and matrix metalloproteinases (MMPs), that specifically degrade extracellular matrix proteins [1-4]

A number of studies have provided evidence of their involvement in remodeling during fibrotic and/or inflammatory processes in the liver and in other organs. MMPs are involved in various processes, including ovulation, embryogenic growth and differentiation, tumor invasion and metastasis [5-8]. The activity of MMPs may be regulated at the level of gene transcription, during proenzyme activation, and by binding of the proenzyme or active enzyme to specific inhibitors (TIMPs). More than 23 MMPs[7] and 4 TIMPs have been identified so far, and, among them, MMP-1,-2,-3,-7,-8,-9,-10,-13,-14,-15 and -16 are expressed in the liver, especially in stellate cells which are well known as main cell source of extracellular matrix, MMPs and TIMPs [9-11]. The intraperitoneal administration of the hepatotoxic CCl_4 to mice, is a well established model to investigate liver fibrosis[12]. Usually, after a single injection of CCl_4, a strong inflammation left the place to liver fibrosis that spontaneously reverts in few days. Upon liver injury, a specific chronological succession chronological succession of events is activated[13,14]. In detail, twelve hours after the intraperitoneal injection of CCl_4 in mice, hepatocyte ballooning began around the central vein, and hepatic necrosis occurrs adjacent to the central vein. Zonal necrosis with infiltrating cells, such as macrophages and lymphocytes, is observed at 48-72 hours around the central vein. The necrotic change is diminished at 5 days with a complete regression to normal liver at 7 days. Reticular fibers are increased around the necrotic areas at 5 days and decreased again at 7 days. This modulation in the ECM is linked to relative increased expression of the MMPs: at 6 hours after the intoxication, the expression of MMP-13 (collagenase 3), which is seldom detected in normal liver, is clearly increased until 12 hours and then decreased again at 24 hours. Others MMPs like the MMP-2 have later peak and gradually decrease at day 7.

Matrix metalloproteinase 13 is a potent zinc-dependent proteolytic enzyme involved in the degradation of the extracellular matrix components such as collagen fibers, gelatin ecc. Its physiologic activity is not restricted to the ECM degradation but it is involved in the processing of cytokines, growth factors and adesion molecules. Besides its physiological role in tissue remodeling, embryogenesis and cell behavior regulation, the proteolytic function, and in

particular the collagenolytic activity of MMP-13 has been demonstrated crucial for the fibrosis resolution in rodent models of liver fibrosis[15]. Indeed, this pathology is characterized by an extensive deposition of type I and type III collagen fibrils synthesized by hepatic stellate cells. At the same time, MMP-13 is believed to promote the early stages of the inflammation that trigger this disease[16]. The already available data on the activity of MMP-13 in rodent liver inflammation and fibrosis was the prompt to better investigate its role in the different steps of the pathology.

Materials and methods

Expression and purification of human MMP13 and MMP12 catalytic domain. The cDNA of proMMP13 (Leu20-Pro268) was cloned into the pET21 vector (Novagen) using *Nde I* and *Xho I* as restriction enzymes. One additional methionine at position 19 was present in the expressed product. The stabilized cat-MMP-13 F175D mutant and the catalytically inactive E223A were produced using the quick-change site-directed mutagenesis kit (Quiagen). The *E. coli* strain BL21 Codon Plus cells, transformed with the above plasmid, were grown in LB medium at 310 K. The protein expression was induced during the exponential growth phase with 0.5 mM IPTG. The cells were harvested 3 hours after induction. After cells lysis, the inclusion bodies, containing proMMP13, were solubilized in 8 M urea; 20 mM Tris (pH 8.0). The protein was purified on the Hitrap Q column (Pharmacia) with a buffer containing 6 M urea; 20 mM Tris (pH 8.0). The elution was performed using a linear gradient of NaCl up to 0.6 M. The purified protein was then refolded using multi-step dialysis against solutions containing 50 mM Tris (pH 7.2); 10 mM $CaCl_2$; 0.1 mM $ZnCl_2$; 0.3 M NaCl and decreasing concentrations of urea. The refolded protein was exchanged, by dialysis, against a buffer with 50 mM Tris (pH 7.2); 5 mM $CaCl_2$; 0.1 mM $ZnCl_2$; 0.3 M NaCl. The protein was concentrated at room temperature using an Amicon concentrator, up to about 30 μM. The active protein is left overnight in these conditions to allow the autoproteolysis of the prodomain. After addition of Acetohydroxamic acid (AHA) to a final concentration of 0.5 M, the catalytic domain of MMP13 (Tyr 104- Pro 268) was purified using size-exclusion chromatography with the buffer 50 mM Tris (pH 7.2); 5 mM $CaCl_2$; 0.1 mM $ZnCl_2$; 0.3 M NaCl, 0.2 M AHA and then concentrated using a Centriprep concentrator at 277 K up to 15 uM.

Cloning, expression and purification of the catalytic domain of MMP12 (F171D mutant) has been previously described. The recombinant proteins were dialyzed against a saline solution

containing 5 mM $CaCl_2$ and 0.1 mM $ZnCl_2$ to prevent protein unfolding. The proteolytic activity of the proteins was assayed by a spectrofluorimeter using the fluorogenic peptide substrate Mca-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH2 (Biomol Inc.).

Animals and protocols. Male, 7 weeks-old, Wistar rats of 250 g body weight, SD1 and SCID/CB17 mice of 25 g body weight were used for the experiments. All animals were fed with Good Laboratory Practice diet in pellets, were housed in plastic cages with a wire-mesh providing isolation from a hygienic bed and were exposed to a 12-hour, controlled light cycle. Experiments were performed in accordance with the institutional ethical guidelines.

CCl_4 was diluted 50% (vol/vol) in olive oil, and the solution was administered in a dose of 4 mL/kg for rats and 1 ml/kg for mice. The animals were randomly divided into four groups: treated with CCl_4 only; with MMP-13 only; with both CCl_4 and MMP-13; and control group, treated with olive oil and physiologic solution.

The animals were killed using diethyl ether anesthesia and terminally bled via cardiac puncture. Kidneys and livers were collected and fixed in 4% phosphate buffered formaldehyde or in methanol for 18–24 hours and embedded in paraffin, for further examinations.

Tissue sections (4μm thick) were stained with H&E staining (Merck) for morphometric analysis or with Sirius Red for collagen staining. The latter were performed by a computerized video-image analysis system.

Results and discussion

Besides the classical collagenolytic activity of MMP13, that has been demonstrated after fibrils deposition[15], the proteolysis of other physiologically relevant substrates by MMP-13 may play a role in the inflammation process that is at the basis of liver fibrosis [16]. The extensive vascularization of the liver together with its very high scavenger activity toward exogenous proteins such as trombolytic enzymes, therapeutic monoclonal antibody, ecc., render the liver a suitable target to evaluate the effect of their intravenous administration. In particular, parenchymal liver cells are responsible for most of the liver up take of the exogenous protein that are then degraded in the lysosomal compartment of these cells[17]. However, to properly evaluate the role of this poorly characterized proteolitic activity of MMP-13 in the pathological process of

liver fibrosis, it is important to design an experimental_procedure to selectively abolish the collagenolityc properties of the protein, without affecting the enzymatic activity toward other substrates.

It has been shown that Hemopexin domain is required for the degradation of the triple helix collagen, while the proteolytic activity toward peptides is usually retained by the isolated catalytic domain[18,19]. Therefore the use of the catalytic domain, instead of the whole enzyme in rodent models of liver inflammation and fibrosis, was selected to provide answers to several open questions about the molecular mechanisms that induce the pathology. In humans, the homologous MMP-1 more than MMP-13 seems to play a critical role in such pathological process. Hence, the human construct, instead of the murine MMP13, was chosen in order to investigate a possible different role of human MMP-13.

At the concentrations required for the i*n vivo* experiments, the *wild type* catalytic domain of MMP13 undergoes to a fast self-proteolysis. Therefore, samples of *wild type* catalytic domain of MMP-13 were investigated by MS-MALDI in order to identify the cleavage sites. Two main fragments arise from the proteolytic cleavage of the protein at the peptidic bond between Asp 174 and Phe 175. Therefore, the protein was stabilized by replacing the lypophilic Phe with one charged aminoacid (Asp) at position 175. This mutation is far from the active site and does not perturb the catalytic activity of the enzyme, as verified by spectrofluorimetric activity assay (data not shown).

Lacking any information about the pharmacokinetic, pharmacodynamic and the *in vivo* activity of MMP-13, the therapeutic doses of thrombolytic enzymes were considered as reference to design the experiments.

The evaluation of acute toxicity of recombinant human cat-MMP13 was carried out in 7 weeks old CD1 mice, weighing about 25 grams. Two doses of MMP13 were tested (1.4 mg/kg and 2.8mg/kg corresponding to 100 and 200 μl of solution injected into the tail vein, respectively). Control mice were treated with 100 μl of physiologic saline solution. The animals were divided in 4 groups: non-treated mice (2 animals); control mice (3 animals) high dose-treated mice (3 animals); low dose-treated mice (3 animals). The animals were injected at day 1 and day 3 and then were sacrificed 24 hours after the last injection. At the time of sacrifice, a macroscopic evaluation of internal organs was carried out. Livers and kidneys were harvested for further analysis. Serial sections of the collected tissues were stained with haematoxylin/eosin and Sirius

Red, for the histological evaluation and for collagen analysis, respectively. The microscopic examination of the stained tissues showed that the administration of cat-MMP13 did not induce any evident alteration of the hepatic parenchyma at macroscopic and microscopic level at both administered doses. On the other hand, kidneys of animals treated with saline solution present a trend of increasing size of the clusters kidney, and an increase of capsules space compared to untreated mice. Same alterations were found in animals treated with cat-MMP13 (2.8mg/kg in 200µl), but the glomerular structure was slightly changed compared to animals treated with 100µl of physiologic solution, that could be explained because the volume of cat-MMP13 was the double of the volume injected into the tail vein of the control mice. Hence, the observed alterations could be due to the volume of injected solution. Long term toxicity of the cat-MMP13 treatment was not evaluated because of the strong possibility of antigenic response of the animals to the MMP-13 administration.

Once the absence of toxic effect of human MMP-13 on mice was evaluated, the effects of the administration of MMP-13 by tail vein injection has been investigated in mouse and rat models of liver fibrosis.

In order to establish the minimum efficacious dose of cat-MMP13 in the rat model, the recombinant protein was administered to a group of rats in a single dose of 0.08, 0.32, 1.4 and 2.8 mg/kg by tail vein injection, 6h after the intraperitoneal administration of 1 ml/kg of CCl_4. The animals were sacrificed 24, 48, 72, and 96 hours after the injection of the CCl_4 and the livers were stained and examined. The histological evaluation of liver specimens from control and treated rats established significant differences related to the administration of MMP-13 and to its dosage. The administration of the protein significantly reduces the score for necrosis and inflammation of liver, speed-up the recovery of liver parenchyma and reduces the infiltrate. The liver-protecting effect is dose-dependent and increase up to 1.4 mg/kg. The administration of a higher dose of MMP-13 (2.8 mg/kg) does not improve the activity. Then doses of 1 mg/kg for CCl_4 and 1.4 mg/kg for cat-MMP13 were selected for animals administration.

First, rats were used as model. Rats, of an average weight of 250 g, were splitted in 4 groups: treated with CCl_4 only, with MMP-13 only, with both CCl_4 and MMP-13; control group was treated with olive oil and physiologic solution. Six hours after the CCl_4 intraperitoneal injection, MMP-13 was administered through the tail vein. All animals were sacrificed 24, 48, 72 hours after MMP13 injection and the liver tissues were harvested. Rats treated only with CCl_4

presented the common pattern of alterations, with ballooning hepatocytes, necrotic areas and strong inflammatory infiltrations. The group treated with both CCl_4 and MMP13, presented a reduced number of degenerating cells at 24 hours after MMP13 injection. At 48 and 72 hours there was as well a decreased alteration of the parenchymal structure of the liver, and there was a markedly reduced infiltration of inflammatory cells. Rats treated only with MMP-13 did not show any alteration in the liver in response to the protein, as previously observed in mice. Data obtained in rat model of liver fibrosis suggest a protective role of MMP13 in CCl_4-induced acute inflammation. This effect is not related to the collagenolytic activity of the enzyme, but it may be due to its proteolytic activity on other substrates in the hepatic tissue.

Following these findings, the effect of cat-MMP13 treatment was evaluated also in mice, that first underwent to a chronic treatment with CCl_4. Males, 7 weeks-old SCID/CB17 mice, with an average body weight of 25g, were divided in 4 groups (6 animals each), as reported for rats. The treatment with CCl_4 was twice a week for 4 weeks, followed by the injection of MMP-13 or saline solution, 6 hours after the last CCl_4 injection. The initial doses and routes of administration tested were of 1 ml/kg of CCl_4 and 1.4 mg/kg of cat-MMP13, but, due to the high mortality after only 1 week of treatment, the experiment was interrupted. The mortality was 33% (2/6) in the control group treated with only CCl_4 and 50% (3/6) in the group treated with CCl_4 and cat-MMP13. At the macroscopic examination, both groups presented large necrotic areas in the liver contest. At the histological analysis, the massive hepatic damage induced by CCl_4 administration did not show differences, compared to the control group. Then, due to the high toxicity of CCl_4 in this mouse line, the CCl_4 concentration was decreased. In the new experimental protocol, 8 animals as control group, and 9 animals as treated group were used. 0.4 ml/kg of CCl_4 was administered twice a week for 4 weeks. In control mice, 100 µl of physiologic saline solution was administered, 6 hours after the last CCl_4 treatment, while in treated mice 1.4 mg/kg of human cat-MMP13 in 100 µl of physiologic solution were injected. The animals were weighted once a week before and after treatments. All animals were sacrificed 24 hours after treatment with MMP13. Then internal organs were macroscopically evaluated, and livers were collected for further biochemical analysis. At the time of explanation, body and liver weights were assessed. After 4 weeks of treatment the weight of the MMP13-treated group was significantly higher than controls (see table 1), but there was no difference in the ratio between the liver and the total body weight.

Table 1 Average body weight of mice, before and after treatment with CCl_4 (A) and with CCl_4 + MMP-13 (B).

	A	B	p value
Before	23±1.14	24,67±1.87	0.06
After	23.96±1.34	26.13±1.53	0.01

The histological analysis of liver from 3 animals of the control group, showed necrotic areas, while none of the animals treated with MMP13 had such alterations. As shown by Sirius Red staining, the administration of cat-MMP13 does not completely block the CCl_4-induced liver fibrotic process, especially in proximity of large vessels, anyway, it was able to reduce the amount of total collagen depositions and the inflammatory process.

To prove that the proteolytic mechanism of MMP13 is responsible for the observed anti-inflammatory activity of cat-MMP13, the recombinant inactive mutant of the protein, obtained by replacing the catalytically relevant glutamate 223 with alanine, was injected into a group of CCl_4-treated animals following the same experimental protocol. The severity of liver inflammation was similar to that observed for the control group which had not received the protein. This experimental evidence supports the hypothesis that the proteolytic activity of the catalytic domain of MMP13 is responsible for the observed fast recovery of liver parenchyma and anti-inflammatory activity.

The specificity of catalytic activity of MMP13 on liver of animals, treated with CCl_4, was verified by the evaluation of the effects of the catalytic domain of MMP12, administered following the same protocol in rats and mice. No protective effect of cat-MMP12 on the CCl_4-induced inflammatory process was observed.

In conclusion, from our studies the acute administration of recombinant human cat-MMP13 is not toxic and does not induce, in mouse and rat models, obvious changes in the liver. Only poorly significative alterations in the kidney are observed, most likely due to the rapid injection of a solution into the tail vein. No data are available on the chronic toxicity, because the proteic nature of the drug involves the risk of an immune response in the animal models.

Data obtained from the treatment of mice and rats models of liver fibrosis induced by CCl4 administration showed that the non-collagenolytic catalytic activity of MMP13 trigger a

protective role during the initial phases of liver fibrosis. In particular, cat-MMP13 plays a key role in the reduction of the inflammatory process, by reducing the amount of necrotic cells and inflammatory cells in the hepatic tissue. Even if our investigations do not reveal the molecular target of the proteolytic activity of MMP13, we can hypothesize that cytokines or adhesion molecules could be involved in this collagenolytic-independent process carried out by MMP13.

Figures and legends:

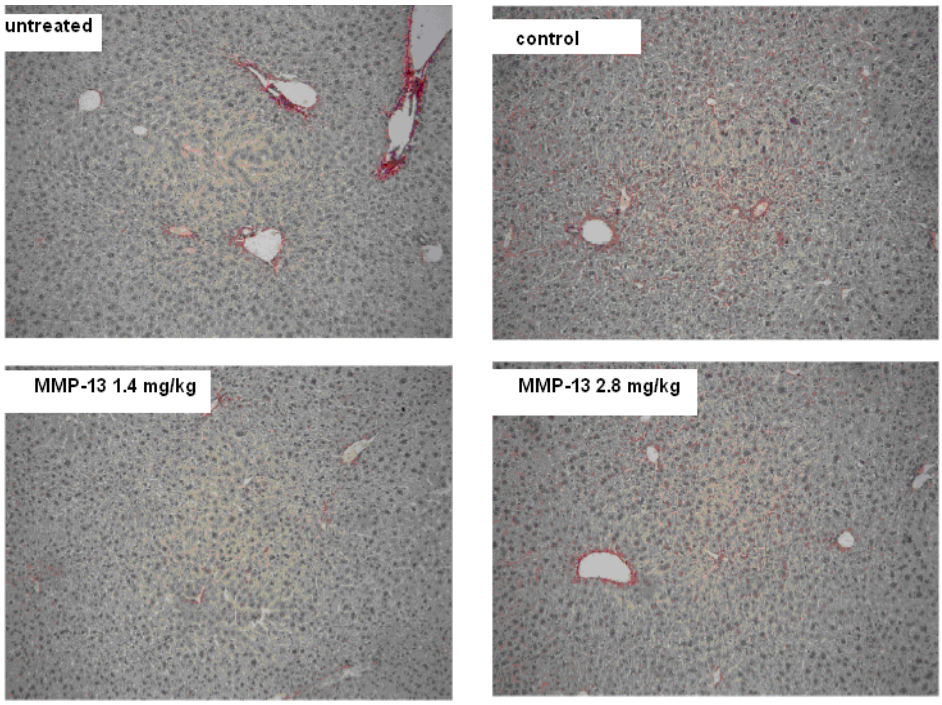


Figure 1 Sirius red-Haematoxylin/eosin staining of livers from CD1 mice.

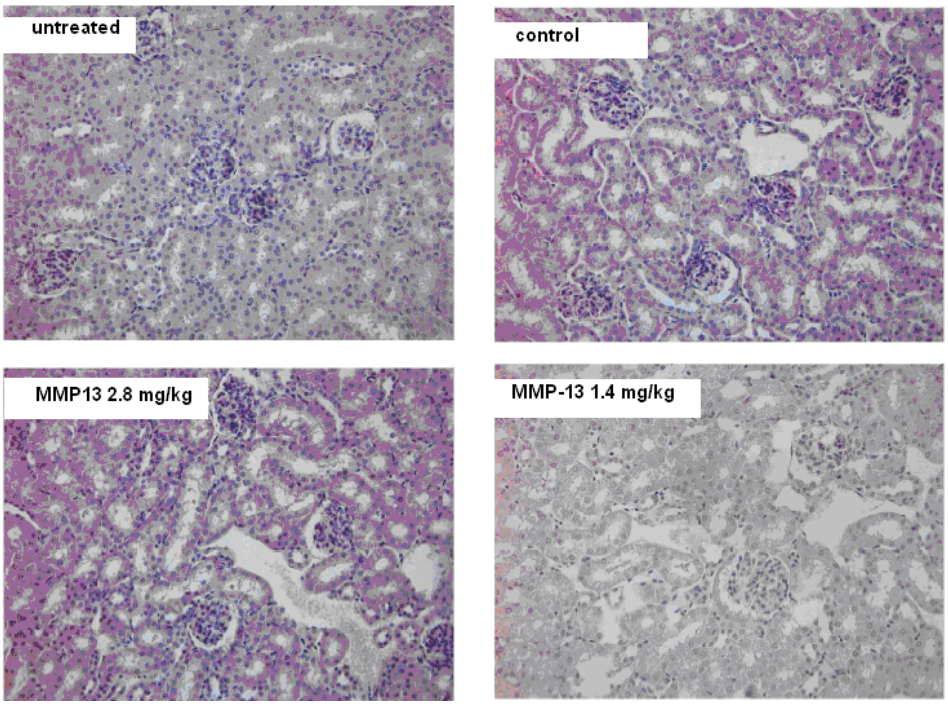


Figure 2 Sirius red-Haematoxylin/eosin staining of kidneys from CD1 mice treated with 1.4 and 2.8 mg/kg of catMMP13.

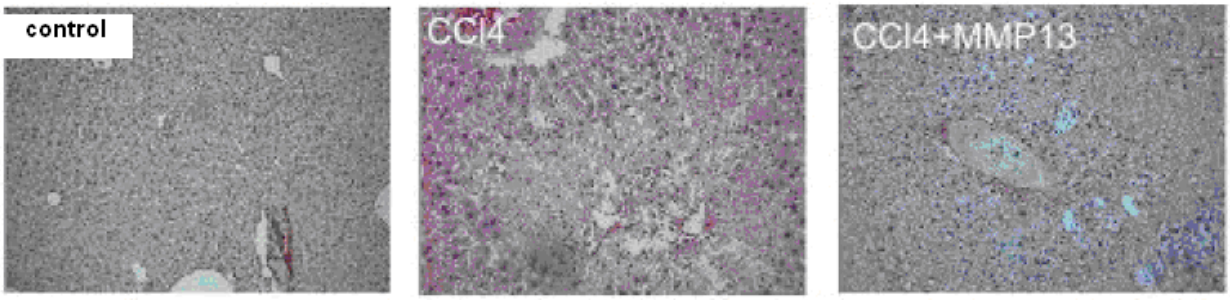


Figure 3 Sirius red-Haematoxylin/eosin staining of livers from SCID mice treated with 1 ml/kg of CCl_4 and 1.4 mg/kg of cat-MMP13.

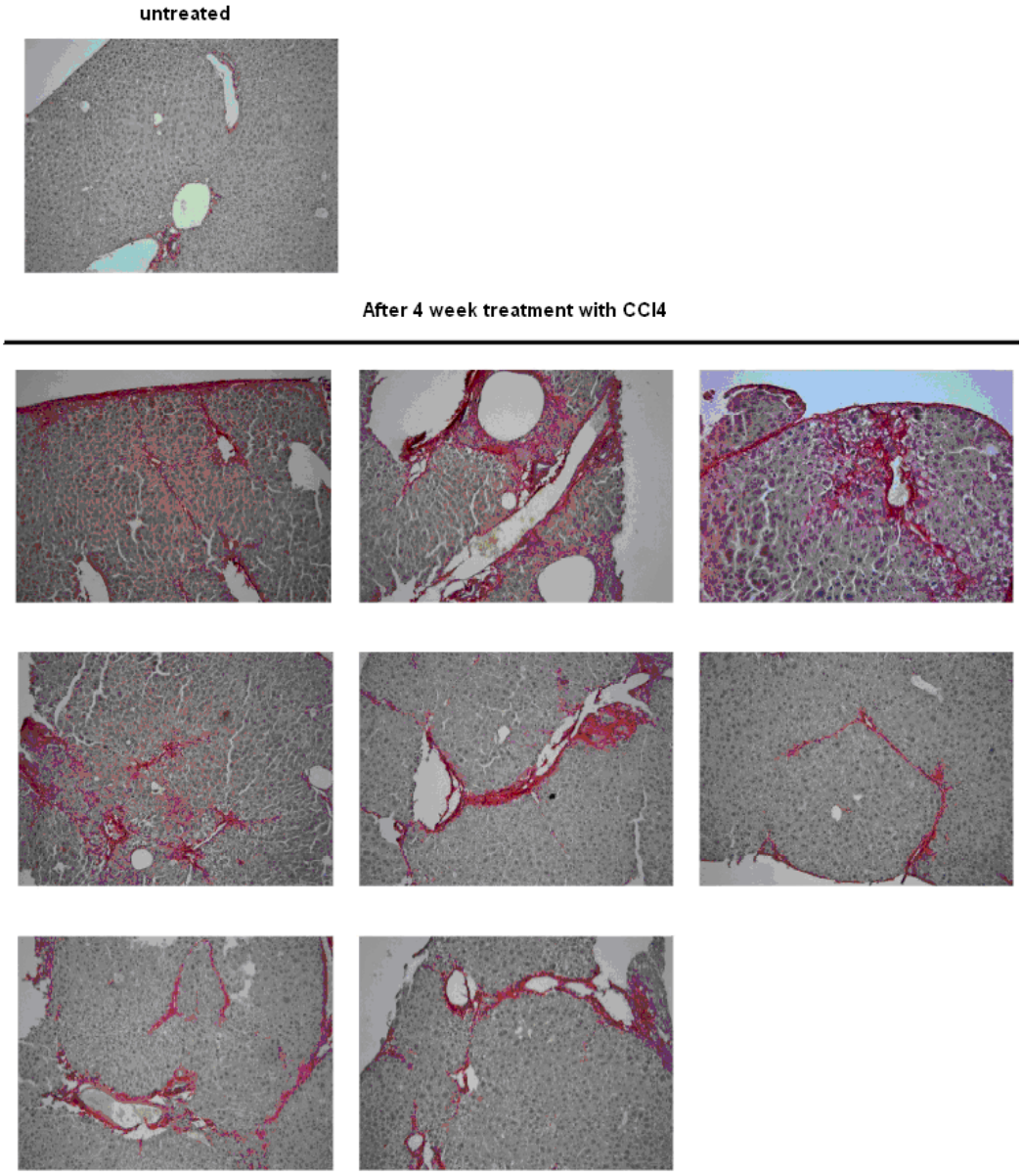


Figure 4 Sirius red-Haematoxylin/eosin staining of livers from SCID mice untreated and treated with 0.4 ml/kg of CCl_4 for 4 weeks.

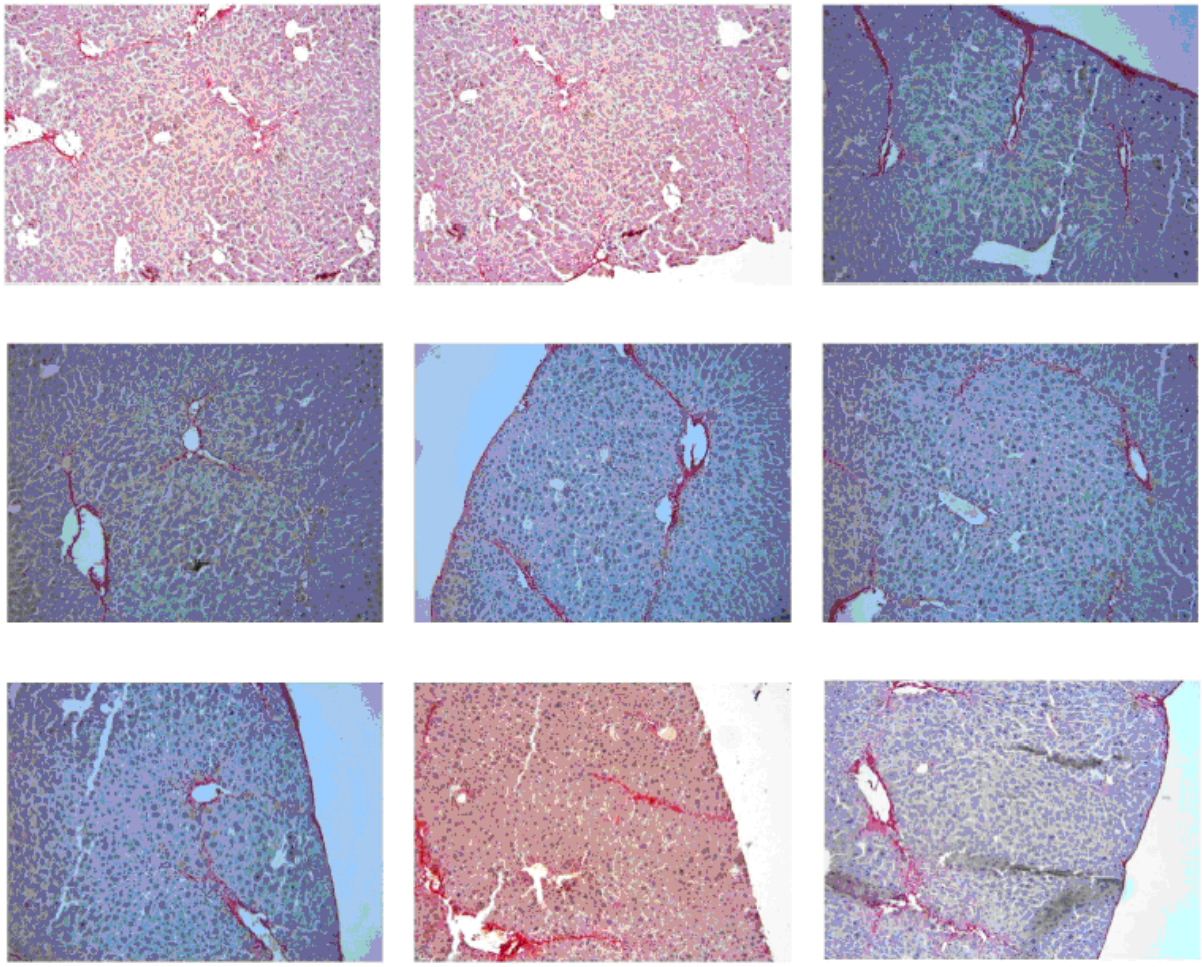


Figure 5 Sirius red-Haematoxylin/eosin staining of livers from SCID mice treated with 0.4 ml/kg of CCl_4 for 4 weeks and then with 1.4 mg/kg of catMMP13.

Reference List

1. Friedman SL. Molecular regulation of hepatic fibrosis, an integrated cellular response to tissue injury. Journal of Biological Chemistry 2000; 275(4):2247-2250.

2. Sarem M. Hepatic stellate cells: their role in normal and pathological conditions. Znaidak R., editor. Gastroenterol.Hepatol. 29, 93-101. 2006. Macias M.

3. Kershenobich Stalnikowitz D. Liver fibrosis and inflammation. A review. Weissbrod A.B., editor. Ann.Hepatol. 2[4], 159-163. 2003.

4. Parsons CJ, Takashima M, Rippe RA. Molecular mechanisms of hepatic fibrogenesis. Journal of Gastroenterology and Hepatology 2007; 22:S79-S84.

5. Sternlicht MD, Werb Z. How matrix metalloproteinases regulate cell behavior. Annu Rev Cell Dev Biol 2001; 17:463-516.

6. Nelson AR, Fingleton B, Rothemberg ML. Matrix metalloproteinases: biologic activity and clinical implications. J Clin Oncol 2000; 18(5):1135-1149.

7. Massova I, Kotra LP, Fridman R, Mobashery S. Matrix Metalloproteinases: structures, evolution, and diversification. FASEB J 1998; 12:1075-1095.

8. Parks WC, Wilson CL, Lopez-Boado YS. Matrix metalloproteinases as modulators of inflammation and innate immunity. Nat Rev Immunol 2004; 4(8):617-629.

9. Schaefer B, Rivas-Estilla AM, Meraz-Cruz N, Arturo M, Reyes R, Hernandez-Nazara ZH et al. Reciprocal modulation of matrix metalloproteinase-13 and type I collagen genes in rat hepatic stellate cells. American Journal of Pathology 2003; 162(6):1771-1780.

10. Knittel T, Mehde M, Grundmann A, Saile B, Scharf JG, Ramadori G. Expression of matrix metalloproteinases and their inhibitors during hepatic tissue repair in the rat. Histochemistry and Cell Biology 2000; 113(6):443-453.

11. Arthur MJP. Fibrogenesis - II. Metalloproteinases and their inhibitors in liver fibrosis. American Journal of Physiology-Gastrointestinal and Liver Physiology 2000; 279(2):G245-G249.

12. Fort J, Oberti F, Pilette C, Veal N, Gallois Y, Douay O et al. Antifibrotic and hemodynamic effects of the early and chronic administration of octreotide in two models of liver fibrosis in rats. Hepatology 1998; 28(6):1525-1531.

13. Takahara Y, Takahashi M, Wagatsuma H, Yokoya F, Zhang QW, Yamaguchi M et al. Gene expression profiles of hepatic cell-type specific marker genes in progression of liver fibrosis. World Journal of Gastroenterology 2006; 12(40):6473-6499.

14. Takahara T, Furui K, Funaki J, Nakayama Y, Itoh H, Miyabayashi C et al. Increased Expression of Matrix Metalloproteinase-Ii in Experimental Liver Fibrosis in Rats. Hepatology 1995; 21(3):787-795.

15. Okazaki I, Watanabe T, Hozawa S, Arai M, Maruyama K. Molecular mechanism of the reversibility of hepatic fibrosis: With special reference to the role of matrix metalloproteinases. Journal of Gastroenterology and Hepatology 2000; 15:D26-D32.

16. Uchinami H, Seki E, Brenner DA, D'Armiento J. Loss of MMP 13 attenuates murine hepatic injury and fibrosis during Cholestasis. Hepatology 2006; 44(2):420-429.

17. Stang E, Krause J, Seydel W, Berg T, Roos N. Endocytosis and Intracellular Processing of Tissue-Type Plasminogen-Activator by Rat-Liver Cells Invivo. Biochemical Journal 1992; 282:841-851.

18. Lauer-Fields JL, Tuzinski KA, Shimokawa K, Nagase H, Fields GB. Hydrolysis of triple-helical collagen peptide models by matrix metalloproteinases. Journal of Biological Chemistry 2000; 275(18):13282-13290.

19. Lauer-Fields JL, Juska D, Fields GB. Matrix Metalloproteinases and collagen catabolism. Biopolymer 66[1], 19-32. 2002.

I. V. S100 Proteins Regulate p53 Oligomerization State

V.1. INTRODUCTION

The p53 protein acts mainly as a transcription factor, although some apoptotic activities are transcriptionally independent. In response to genotoxic stress, p53 translocates in the nucleus where it can form the active homotetramer and then activate the expression of genes necessary to maintain genomic stability[1-4].

The activity of p53 is regulated at different levels, including post-translational modifications, p53 stability, localization and tetramerization[5-12]. Tetramerization is relevant for p53 activity at three cross-regulated levels: stability, cellular localization and DNA binding. Indeed, p53 nuclear translocation is more effective for the monomeric form, moreover p53 tetramerization cover a NES signal, leading to nuclear retention. Localization is important also in the regulation of p53 stability, indeed nuclear p53 is ubiquitinated by MDM2 and ubiquitination contributes to its export in the cytoplasm, where it is degraded by a MDM2-mediated mechanism[13]. Finally, the binding of p53 to DNA is cooperative, with the dimer having less affinity for the DNA with respect to the tetramer, which is the active form of the transcription factor[14].

In more than 50% of all human cancers, the loss of p53 activity by mutations is a molecular key event in deregulation of cellular homeostasis and in the development of tumors[15]. Anyway, in a lot of cancer cells expressing *wild type* p53, inhibition of tumor suppressor activity has been found to be correlated with deregulation of proteins involved in modulation of p53 activity.

Some members of the S100 family interact with p53, and they exert different effects on p53 activity[6]. The most studied are S100A4 and S100B, which are thought to inhibit the phosphorylation of p53 C-terminal region, and p53 tetramerization, leading to inhibition of its transcriptional activity, thereby compromising p53 tumour-suppressor activity[16-19]. In contrast, S100A2 promotes p53 transcriptional activity[20] and interestingly S100A4 has also been documented as enhancing p53-dependent apoptosis[16]. Thus the balance of actions of different S100 proteins within a cell will determine the function of p53.

The binding properties of some S100 proteins with peptides derived from p53 has been characterized through different techniques, and these data suggest that different S100 proteins bind p53 in different ways, anyway only the p53-S100B complex has been characterised from

a structural point of view[21-24]. The binding of S100 proteins to p53 is dependent upon p53 oligomerization state, post-translational modifications and peptide length. Moreover, S100 proteins can bind different regions of p53 (NRD, TET domain, NLS), in particular, it has been proposed that S100s interacting with p53 NRD have a negatively charged hinge region (residues 45-51), while those that do not bind have a positive or null charge[25].
These binding properties could explain the observed different effects in the modulation of p53 activity by some S100s, moreover the investigation of the interaction between S100 proteins and full length p53 could lead to some clarifications about the role of S100 proteins in the modulation of p53 activity.

V.2. MATERIAL AND METHODS

P53 cDNA source. The human colorectal adenocarcinoma cell line SW620 (ATCC), expressing p53(R273H, P309S) mutant, was maintained in Dulbecco's modified Eagle's medium supplemented with 100 U/ml penicillin and 100 μg/ml streptomycin and 10% foetal bovine serum in a humidified 5% CO_2:95% air incubator at 37°. Total RNA was extracted from confluent cells with RNeasy Mini Kit (Qiagen), retrotranscribed with Omniscript Reverse Transcription kit (Qiagen) and cDNA was used as template for p53 cDNA amplification.

Cloning. Full length-, Δ93-, Δ292-p53 constructs were amplified by PCR from SW620 cDNA, using specific primers for Gateway Cloning System (Invitrogen) and cloned in different pDEST vectors, fused with N-terminal tags: 6xHis, GST, NusA, TrxA, MBP. P53Δ310 was cloned in pET15b vector (Novagen) in frame with the N-terminal 6xHis-tag, using NdeI and BamHI restriction sites.

Mutagenesis. The *wild type* sequence (NM 000546) was obtained by H273R, S309P retromutations, and the M133L, V203A, N239Y, N268D mutations were inserted in the *wild type* sequence of full length- and Δ93-p53 to obtain the superstable mutant[26]. Mutagenesis reactions were performed with specific primers using the Quick Change Site-directed Mutagenesis kit (Stratagene). All sequences were verified. The full length p53 contains the P72R SNP (Ref SNP ID: rs1042522).

Recombinant proteins expression. All the fusion proteins were expressed in inclusion bodies in any tested conditions. The full length- and Δ93-p53 His tag fusion proteins were expressed in C41(DE3) E. coli strain. P53Δ292 and p53Δ310 were expressed in BL21-gold E. coli strain as His tag fusion proteins. Expression conditions were optimised for each construct. The proteins were extracted from inclusion bodies, purified by HisTag affinity chromatography and refolded by dialysis. The refolded proteins were further purified by size exclusion chromatography. The protein folding was verified by 1D NMR spectroscopy.

Light Scattering. The oligomerization state of concentrated samples of full length p53 was analysed by size exclusion chromatography and light scattering (Wyatt Technology). 100 μl of samples containing increasing concentrations of full length p53 were loaded in a $G2000SW_{XL}$ analytical column (Tosoh Bioscience) with a void volume of 150 kDa. The column was connected to a multi-angle light scattering spectrometer. Data were analysed with Astra software (Wyatt Technology), to estimate the molar mass.

NMR. ^{1}H-^{15}N HSQC and ^{1}H-^{15}N TROSY-HSQC spectra of ^{15}N labelled S100A2 and S100P were acquired at 600 and 800 MHz, with and without unlabelled p53, either in presence and in absence of Ca^{2+} in 25mM Hepes, 150mM NaCl, 2mM DTT, pH 7.5.

S100P-p53 interaction. His tagged full length p53 was immobilised on a His Gravitrap column (GE Healthcare) in 25mM Hepes, 150mM NaCl, 2mM DTT, pH 7.5, with and without 10 mM $CaCl_2$. Untagged S100P was loaded on the column with and without immobilised p53, either in presence or absence of $CaCl_2$.

V.3 RESULTS AND DISCUSSION

P53 expression. The DNA sequence of p53 has been amplified using as template the cDNA retrotranscribed from the RNA extracted from a colon cancer cell line, SW620. This cell line has been chosen because normal cells have very low levels of p53 RNA and the amplification of p53 cDNA by PCR is impaired. On the contrary, SW620 cell line express high levels of p53(R273H, P309S) mutant and the *wild type* sequence can be easily obtained by retromutations. The recombinant expression of native full length p53 in E. coli is very difficult, because of the high molecular weight (180 kDa as tetramer), the presence of large unstructured regions both in the N-terminal and C-terminal of p53, the large number of Cys

residues, and the aggregation propensity. All the expression tests were conducted as reported in Table 1.

All the constructs were expressed in inclusion bodies in any tested conditions, then only the His tag fusion protein was carried out for refolding. The yield of purified protein for the His tagged construct is 45-50 mg/L in LB growth medium. Refolding was carried out with a first dilution of denaturing agent (urea or Gnd-HCl), then with eight steps dialysis to obtain the folded protein in 25 mM Sodium Phosphate buffer (pH 7.2), 150 mM NaCl, $ZnCl_2$ 0.1 mM, and 5 mM DTT. Zn^{2+} ions in the DBD are necessary for thermodynamic stability and reduce aggregation. The protein can be concentrated up to 0.2 mM, higher concentrations induce protein precipitation.

Table 1.

EXPRESSION TESTS PARAMETERS	
Fusion tags	HisTag / TrxA-HisTag / GST-HisTag / NusA-HisTag
E. Coli strains	BL21Gold – Origami – Rosetta pLysS C41(DE3) – BL21(DE3)Codon Plus RIPL
Induction temperature (°C)	37 – 30 – 25 – 18
IPTG concentrations (mM)	1 – 0.7 – 0.5 – 0.1
Growth medium:	LB – 2xYT – minimal medium (M9)

The preparation of ^{15}N labelled p53 was not possible, since in any tested conditions the protein was expressed but it was unable to refold. This problem in the preparation of labelled samples, impaired the possibility to carry out NMR experiments for the structural characterizations of FL-p53 in complex with S100 proteins.

P53 aggregation was evaluated at different protein concentrations and buffer conditions, by size exclusion chromatography and multi-angle light scattering. P53 samples were loaded in a $G4000SW_{XL}$ analytical column (Tosoh Bioscience), with a void volume of 7000 kDa. The column was connected to a multi-angle light scattering spectrometer. The chromatogram in Fig. 1 shows the profile of UV absorbance and scattered light, obtained with a 200 μM sample of full length p53 in 25 mM HEPES, 150 mM NaCl, 2 mM DTT, pH 7.5. For higher concentration the amount of aggregates was higher and protein precipitation is induced.

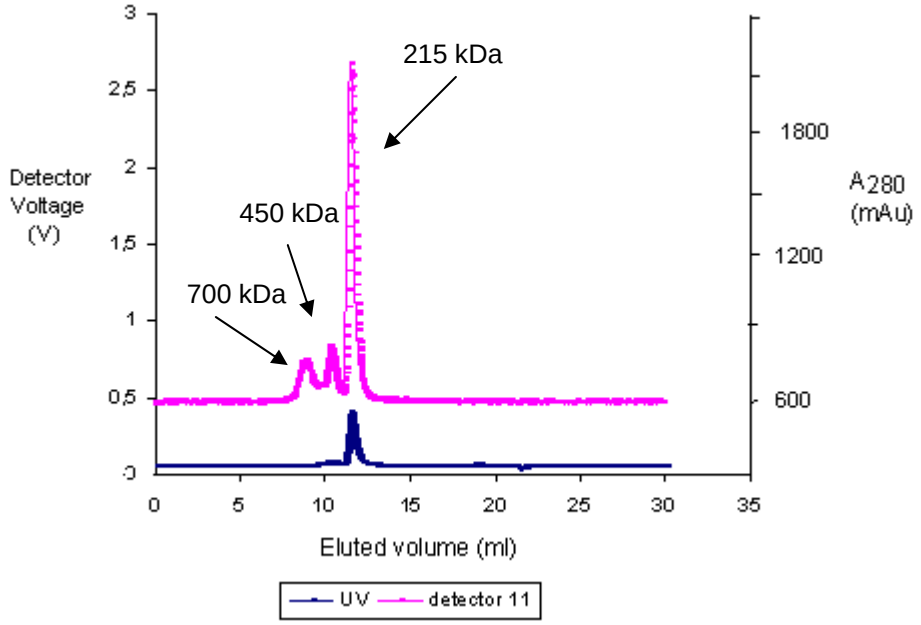


Fig. 1. Size exclusion and light scattering analysis of full length p53 aggregation.

The interactions of S100 proteins with the pseudo *wild type* full length p53 were carried out by NMR spectroscopy using ^{15}N S100 protein and unlabeled p53.

P53-Apo-S100A2 interaction. Equimolar amounts of fl-p53 and apo-S100A2 were mixed at 1 μM concentration in 25 mM HEPES, 150 mM NaCl, 2 mM DTT, pH 7.5, in order to prevent proteins precipitation. The sample was concentrated up to 0.1 mM.

The sample of Ca^{2+}-loaded S2100A2 and p53 was prepared by mixing apo proteins in diluted conditions and adding a concentrated solution of $CaCl_2$ up to 1 mM, to obtain the holo sample. A partial precipitation occurred during concentration in presence of $CaCl_2$, and the samples were clarified by centrifugation.

As shown in Fig.2A, the ^{1}H-^{15}N HSQC spectrum of apo-^{15}N S100A2 at 298 K was not affected by addiction of equimolar amount of p53, while in presence of $CaCl_2$ (Fig. 2B) the ^{1}H-^{15}N HSQC spectrum of ^{15}N S100A2 in presence of p53 was not detectable at 500 MHz, and only few chemical shifts were detected performing a CRINEPT-TROSY-^{1}H-^{15}N HSQC at higher temperature (308 K) with a 800 MHz spectrometer (Fig. 2C).

These experiments confirmed that the interaction of S100A2 with p53 is Ca^{2+}-dependent, but the quality of our acquisitions in these experimental conditions does not allow further structural characterizations. This can be due to partial proteins precipitation after mixing, that decrease the concentration of the sample, and to the formation of a high molecular weight

complex. Also the concentration of the samples (around 0.1 mM) impaired the acquisition of NMR spectra of the complex. Higher concentrations did not improved the quality of the acquired NMR spectra because of proteins aggregation.

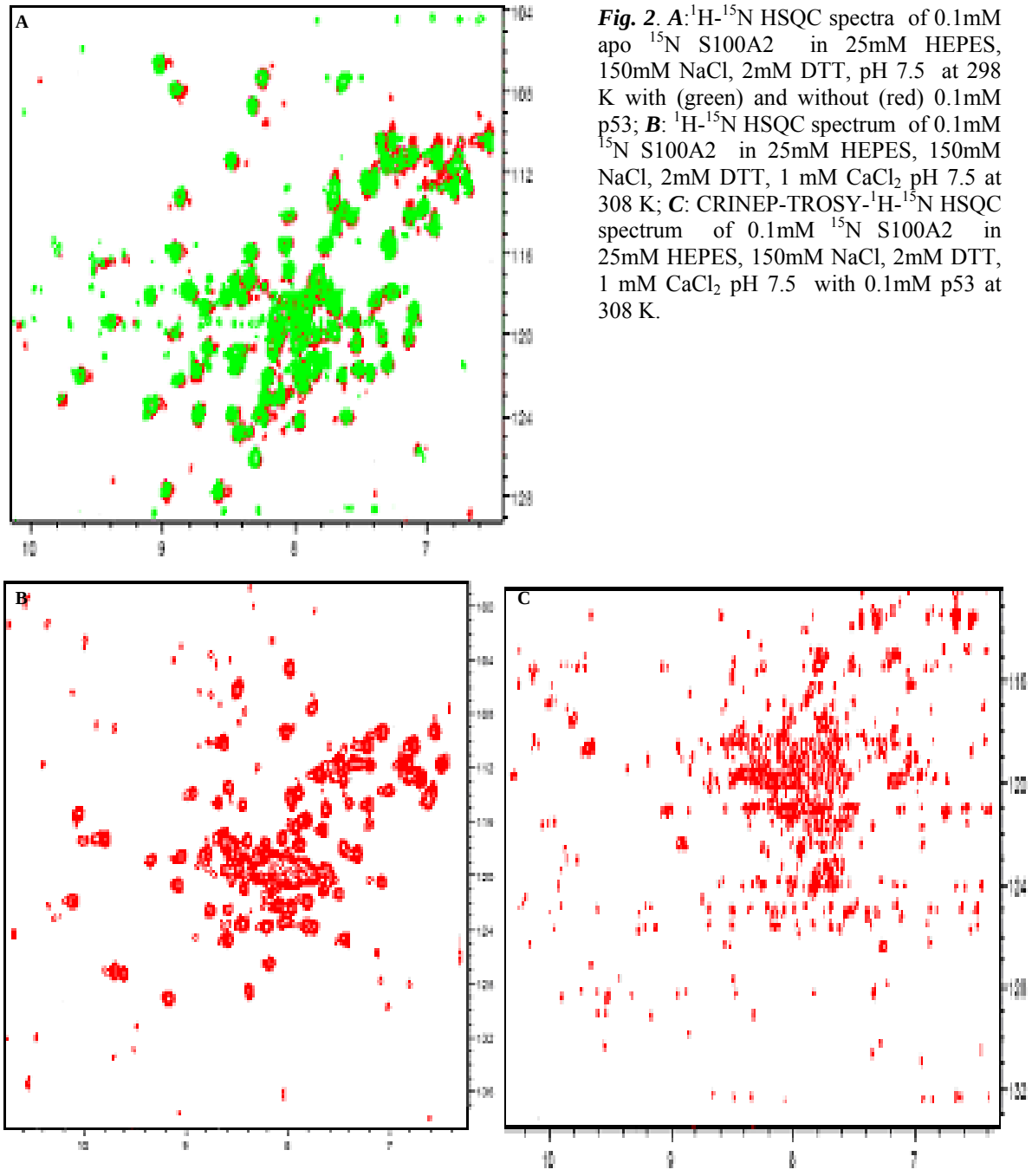


Fig. 2. ***A***:^{1}H-^{15}N HSQC spectra of 0.1mM apo ^{15}N S100A2 in 25mM HEPES, 150mM NaCl, 2mM DTT, pH 7.5 at 298 K with (green) and without (red) 0.1mM p53; ***B***: ^{1}H-^{15}N HSQC spectrum of 0.1mM ^{15}N S100A2 in 25mM HEPES, 150mM NaCl, 2mM DTT, 1 mM $CaCl_2$ pH 7.5 at 308 K; ***C***: CRINEP-TROSY-^{1}H-^{15}N HSQC spectrum of 0.1mM ^{15}N S100A2 in 25mM HEPES, 150mM NaCl, 2mM DTT, 1 mM $CaCl_2$ pH 7.5 with 0.1mM p53 at 308 K.

The interaction between p53 and S100A2 was observed also by agarose native gel electrophoresis. Samples were prepared as described for NMR experiments, and were analysed by native agarose gel electrophoresis. 1% agarose gel was prepared in buffer A (20 mM HEPES, 19.2 mM Glycine, DTT 5 mM, pH 7.4) and the comb placed in the center of the

gel. The gel was submerged in a reservoir containing Buffer A and electrophoresis was performed at a constant voltage of 50 V for 2 h at 4°C. As shown in Fig.3A, when p53 and S100A2 are mixed in presence of $CaCl_2$, in the native gel a third band can be detected, which migrate toward the cathode. This band has been extracted from gel and loaded in SDS-PAGE, to demonstrate that it correspond to the complex between p53 and S100A2 (Fig. 3B).

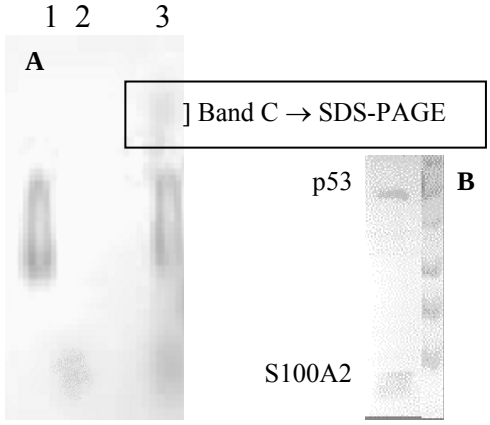


Fig. 3. Analysis of complex formation of p53 and S100A2 by native gel electrophoresis and SDS-PAGE.
S100A2 and p53 were incubated at a 1:1 molar ratio in 25 mM Hepes, 150 mM NaCl, 2 mM DTT, pH 7.5 for 1 hour at 4°C and centrifuged 5 min at 14,000 *g*. The supernatants were loaded onto a 1 % native agarose gel (A) prepared in 20 mM HEPES, 19.2 mM Glycine, DTT 5 mM, pH 7.4 . Lane 1, p53; lane 2, S100A2; lane 3, S100A2 and p53 (1:1 molar ratio). Band C was extracted from gel and loaded in 15 % SDS-PAGE (B).

P53-S100P interaction. S100P is a member of S100 proteins, which has never been reported to interact with p53. The sequence analysis of S100P, aligned with S100B and S100A4, which are reported to interact with p53, highlights slightly negative charge of loop 2 at physiological pH. Also loop 2 of S100A4, which is reported to interact only with p53 TET domain, has the same charge at physiological pH. On the contrary, in S100B, like S100A1 and S100A2, which are reported to interact with p53 NRD and TET domains, loop 2 has a net negative charge a physiological pH (Fig. 4). These data suggest that S100P should interact with p53 TET domain, but not with NRD.

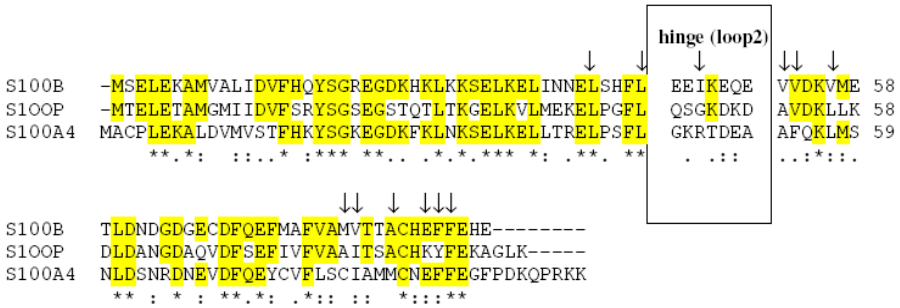


Fig. 4. Aminoacid sequences alignment of S100P with S100B and S100A4, with in evidence the residues of hinge (loop 2). Arrows indicate the residues of S100B involved in binding with p53 NRD as reported by Weber et al.

The interaction of S100P with FL-p53 was verified by agarose native gel, but the band corresponding to the complex cannot be detected in any tested conditions.

Therefore, His GraviTrap column was used to evaluate the ability of untagged S100P to bind His tagged p53 immobilised on the column. The experiment were performed in 25 mM

HEPES, 150 mM NaCl, 2 mM DTT, pH 7.5 (buffer C) either in presence and in absence of 10 mM $CaCl_2$. Tagged p53 was immobilised in the column, loading the protein without Imidazol, and unbound protein was washed out with 40 mM Imidazol. Then untagged S100P was loaded in that column and its retention in the column was evaluated by washing with buffer C + 40 mM Imidazol, and eluting the retained proteins with buffer C + 0.5 M Imidazol.

As shown in Fig. 4, S100P is retained in the column only in presence of $CaCl_2$ (Fig. 5A and B). In order to exclude possible interactions between the column and S100P in presence of $CaCl_2$ and to demonstrate the Ca^{2+}-dependent interaction with immobilised p53, S100P was loaded onto His GraviTrap column without immobilised p53, in presence of 10 mM $CaCl_2$ (Fig. 5C). These experiments confirmed that S100P can bind p53 in presence of $CaCl_2$.

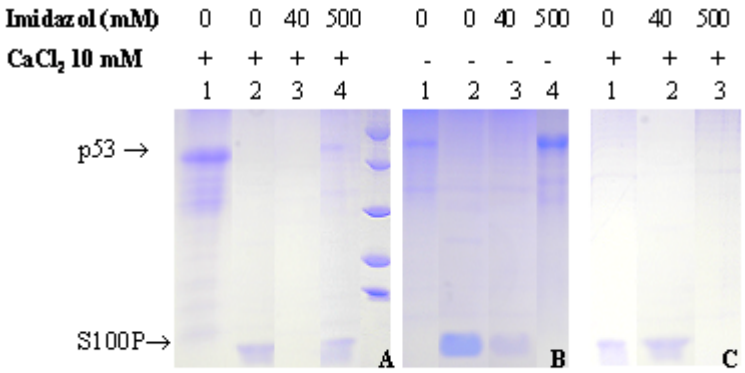


Fig 5 : P53-S100P interaction in His graviTrap column.
A: Lane 1, loaded p53; Lane 2, loaded S100P; Lane 3, elution.
B: Lane 1, loaded p53; Lane 2,loaded S100P, Lane 3, washing; Lane 4, Elution.
C: Lane 1, loaded S100P, Lane 2, washing; Lane 4 Elution.

The interaction between p53 and S100P was evaluated also by NMR spectroscopy (Fig. 6). NMR samples were prepared as described for the interaction with S100A2. ^{1}H-^{15}N HSQC spectra of ^{15}N S100P were performed with and without p53, in presence of 10 mM $CaCl_2$ at 700 MHz. In presence of p53, all peaks of ^{15}N S100P disappeared, indicating the formation of higher molecular weight complex.

These results prove that also S100P can bind p53 with a Ca^{2+}-dependent mechanism and like the other S100 proteins could be involved in the regulation of p53 activity.

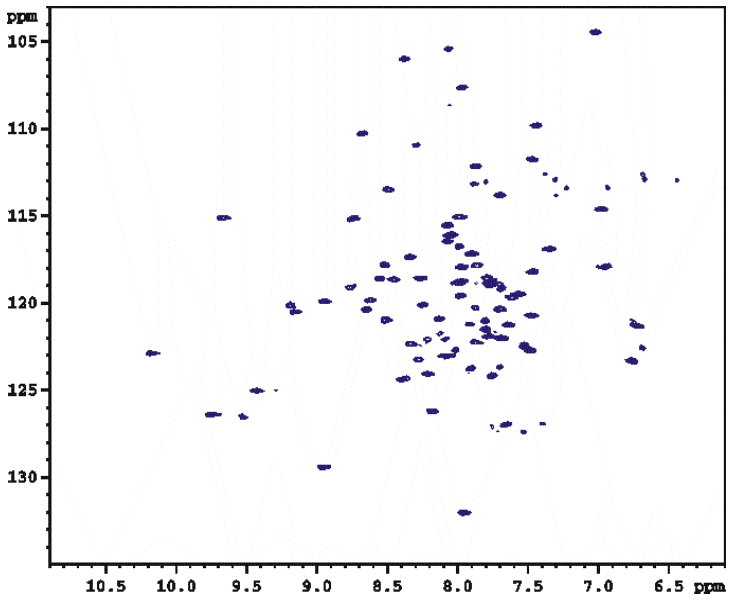


Fig. 6. ^{1}H-^{15}N HSQC spectrum of 0.1 mM ^{15}N S100P in 25 mM HEPES, 150 mM NaCl, 2 mM DTT, 10 mM $CaCl_2$, pH 7.5. The ^{1}H-^{15}N HSQC spectrum of 0.1 mM ^{15}N S100P after addiction of equimolar amount of unlabeled p53 acquired in the same conditions was empty.

V.4. PERSPECTIVES

Structural information on S100s-p53 interactions will be acquired by NMR spectroscopy on the complex of S100 proteins with p53 C-terminal constructs (p53Δ292 and Δ310) in order to reduce the molecular weight of the complex and improve the NMR data.

The superstable mutant full length p53 will be cloned in a modified a pRSETa plasmid with a N-terminally His-tagged lipoyl domain from Bacillus stearothermophilus dihydrolipoamide acetyltransferase (residues 1-87) which has been cloned by M. Allen and co-workers and allow the expression of soluble p53.

Reference List

1. Levine AJ, Hu W, Feng Z. The P53 pathway: what questions remain to be explored? Cell Death and Differentiation 2006; 13(6):1027-1036.

2. Balint E, Vousden KH. Activation and activities of the p53 tumour suppressor protein. British Journal of Cancer 2001; 85(12):1813-1823.

3. Vogelstein B, Lane D, Levine AJ. Surfing the p53 network. Nature 2000; 408(6810):307-310.

4. Bourdon JC. p53 and its isoforms in cancer. British Journal of Cancer 2007; 97(3):277-282.

5. Bode AM, Dong ZG. Post-translational modification of p53 in tumorigenesis. Nature Reviews Cancer 2004; 4(10):793-805.

6. Fernandez-Fernandez MR, Veprintsev DB, Fersht AR. Proteins of the S100 family regulate the oligomerization of p53 tumor suppressor. Proc Natl Acad Sci U S A 2005; 102(13):4735-4740.

7. Lu H, Levine AJ. Human Taf(Ii)31 Protein Is A Transcriptional Coactivator of the P53 Protein. Proceedings of the National Academy of Sciences of the United States of America 1995; 92(11):5154-5158.

8. Marine JC, Jochemsen AG. Mdmx as an essential regulator of p53 activity. Biochemical and Biophysical Research Communications 2005; 331(3):750-760.

9. Nie LH, Sasaki M, Maki CG. Regulation of p53 nuclear export through sequential changes in conformation and ubiquitination. Journal of Biological Chemistry 2007; 282(19):14616-14625.

10. Scotto C, Delphin C, Deloulme JC, Baudier J. Concerted regulation of wild-type p53 nuclear accumulation and activation by S100B and calcium-dependent protein kinase C. Mol Cell Biol 1999; 19(10):7168-7180.

11. Scoumanne A, Chen X. Protein methylation: a new mechanism of p53 tumor suppressor regulation. Histology and Histopathology 2008; 23(9):1143-1149.

12. Vassilev L, Vu B, Graves B, Carvajal D, Podlaski F, Filipovic Z et al. - In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. Science 2004; 303(5659):844-848.

13. O'Keefe K, Li HP, Zhang YP. Nucleocytoplasmic shuttling of p53 is essential for MDM2-mediated cytoplasmic degradation but not ubiquitination. Molecular and Cellular Biology 2003; 23(18):6396-6405.

14. Weinberg RL, Veprintsev DB, Fersht AR. Cooperative binding of tetrameric p53 to DNA. Journal of Molecular Biology 2004; 341(5):1145-1159.

15. Hollstein M, Sidransky D, Vogelstein B, Harris CC. P53 Mutations in Human Cancers. Science 1991; 253(5015):49-53.

16. Grigorian M, Andresen S, Tulchinsky E, Kriajevska M, Carlberg C, Kruse C et al. Tumor suppressor p53 protein is a new target for the metastasis-associated Mts1/S100A4 protein - Functional consequences of their interaction. Journal of Biological Chemistry 2001; 276(25):22699-22708.

17. Lin J, Yang Q, Yan Z, Markowitz J, Wilder PT, Carrier F et al. Inhibiting S100B restores p53 levels in primary malignant melanoma cancer cells. J Biol Chem 2004; 279(32):34071-34077.

18. Wilder PT, Lin J, Bair CL, Charpentier TH, Yang D, Liriano M et al. Recognition of the tumor suppressor protein p53 and other protein targets by the calcium-binding protein S100B. Biochim Biophys Acta 2006; 1763(11):1284-1297.

19. Lin J, Blake M, Tang C, Zimmer D, Rustandi RR, Weber DJ et al. Inhibition of p53 transcriptional activity by the S100B calcium-binding protein. J Biol Chem 2001; 276(37):35037-35041.

20. Mueller A, Schafer BW, Ferrari S, Weibel M, Makek M, Hochli M et al. The calcium-binding protein S100A2 interacts with p53 and modulates its transcriptional activity. Journal of Biological Chemistry 2005; 280(32):29186-29193.

21. Delphin C, Ronjat M, Deloulme JC, Garin G, Debussche L, Higashimoto Y et al. Calcium-dependent interaction of S100B with the C-terminal domain of the tumor suppressor p53. J Biol Chem 1999; 274(15):10539-10544.

22. Rustandi RR, Baldisseri DM, Drohat AC, Weber DJ. Structural changes in the C-terminus of Ca2+-bound rat S100B (beta beta) upon binding to a peptide derived from the C-terminal regulatory domain of p53. Protein Sci 1999; 8(9):1743-1751.

23. Rustandi RR, Baldisseri DM, Weber DJ. Structure of the negative regulatory domain of p53 bound to S100B(betabeta). Nat Struct Biol 2000; 7(7):570-574.

24. Rustandi RR, Drohat AC, Baldisseri DM, Wilder PT, Weber DJ. The Ca(2+)-dependent interaction of S100B(beta beta) with a peptide derived from p53. Biochemistry 1998; 37(7):1951-1960.

25. Fernandez-Fernandez MR, Rutherford TJ, Fersht AR. Members of the S100 family bind p53 in two distinct ways. Protein Sci 2008; 17(10):1663-1670.

26. Nikolova PV, Henckel J, Lane DP, Fersht AR. Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. Proceedings of the National Academy of Sciences of the United States of America 1998; 95(25):14675-14680.

VI. Characterization of CaM targets with unknown structures and binding properties

VI.1. Introduction

Calmodulin (CaM) is a CALcium MODULated proteIN, widely expressed in the cytoplasm of all higher eukaryotic cells and has been highly conserved through evolution. Calmodulin transduces signals to enzymes, ion channels and other proteins in response to variations of intracellular calcium levels. Calmodulin's target proteins come in various shapes, sizes and sequences and are involved in a lot of signal transduction pathways[1-9].

Tuberin is a 198 kDa tumor suppressor protein containing 1807 amino acids and codified by tuberous sclerosis 2 gene(TSC2). Mutations in the TSC2 gene have been genetically linked to the pathology of both tuberous clerosis disease (TSC) and lymphangioleiomyomatosis (LAM), classified as disorders of cellular migration, proliferation, and differentiation[10-12]. The C terminal domain of tuberin has been reported to have GAP activity on Rheb small GTPase, involved in the mTOR signalling[13-17], moreover it contains a CaM binding domain overlapped with the Estrogen Receptor α (ERα) binding domain. Moreover, in this domain a nuclear localization signal has been found to overlap to CaM/ERα binding domain[18,19]. Deletion mutagenesis studies suggested that mutations in CaM binding domain may be involved in the pathology of TSC and LAM.

The CaM binding domains of most CaM target proteins are mainly random coil in solution, but adopt α-helical structures in the complex with CaM.

A comparison of tuberin CaM binding domain with known CaM target peptides, has shown three conserved hydrophobic residues (W1740, L1744, I1747), aligned with the CaM binding peptide of plasma membrane Ca^{2+} pump. Unique among CaM targets, this Ca^{2+} pump can be activated by the C-terminal but not the N-terminal half of CaM[20], suggesting a similar binding between CaM and tuberin. As CaM interacts with a variety of different target enzymes, the determination of structural differences between different CaM complexes is of great importance for the understanding of molecular recognition and specific signaling pathways.

The interaction between CaM and tuberin has been studied in solution NMR, in order to obtain structural information on the complex. CaM N60D can selectively bind Ln^{3+} ions in place of Ca^{2+} in the second Ca^{2+}-binding site of the N-terminal domain [21]. Ln^{3+} ions were used to align the CaM-tuberin complex in a magnetic field and obtain information about the

flexibility of the two domains of CaM in complex with tuberin, through analysis of NH residual dipolar couplings (RDCs). Indeed, if in the complex the CaM domains are blocked, the RDC values are the same for N- and C-terminal domain, but if the two domain are flexible the RDC values for the C-terminal domain are smaller than those of the N-terminal domain.

VI.2. Material and Methods

CaM Targets screening. A list of six interesting human Calmodulin interacting proteins was obtained by searching in different databases. The coding sequences of selected CaM targets were cloned into different Gateway expression vectors and screened for expression and solubility in E. coli strains. Refolding trials were carried out for insoluble proteins.

Tuberin expression. Tuberin (1531-1758) expression vector was transformed in BL21 (DE3) Gold E. coli strain (Novagen). Bacteria were grown at 37°C until OD_{600} of 0.8 was reached. The protein expression was induced by addiction of 0.5 mM IPTG. The cells were allowed to growth further at 18°C for 18-20 hours and then harvested by centrifugation. Soluble proteins were extracted by sonication in ice-cold lysis buffer (20 mM HEPES pH 7.4, 0.2 M KCl, 10 mM Imidazol, 1 mM TCEP , 1 mM Pefabloc). Clarified soluble extract was loaded onto Ni^{2+}-loaded HiTrap Chelating column, and tuberin was eluted in 20 mM HEPES pH 7.4, 0.2 M KCl, 0.5 M Imidazol, 1 mM TCEP, 1 mM Pefabloc, 10 mM EDTA. Purified tuberin was concentrated and further purified by size exclusion chromatography. All purification steps were carried out at 4°C, and protein was never concentrated more than 0.4 mM to avoid aggregation.

Mutagenesis. A stop codon was inserted after the codon of Ile1754, to remove the last four aminoacids (CEEA) of tuberin (1531-1758), using QuickChange Mutagenesis kit (Stratagene).

Tuberin peptide. Synthetic peptide of Tuberin (1740-1754), corresponding to the CaM binding domain of tuberin was obtained from Inbios s.r.l. The lyophilised powder was resuspensed in 20 mM HEPES pH 7.4, 0.15 M KCl, 1 mM TCEP at the concentration of 10 mg/ml and used for titration experiments with N60D CaM.

Samples preparation. ^{15}N apo CaM N60D 1 mM in 20 mM HEPES, 150mM KCl, 1 mM TCEP, pH 7.4, was titrated with $CaCl_2$ solution to obtain the Ca_4CaM form. The Ca_3LnCaM N60D was obtained by titration with $CaCl_2$ and $LnCl_3$ solutions (Ln^{3+}=Tb^{3+}, Yb^{3+}, Tm^{3+}, Dy^{3+}). Titration were followed by ^{1}H-^{15}N HSQC spectroscopy.

NMR. Titrations of ^{15}N Ca_4CaM N60D with unlabeled tuberin (1531-1754) and (1740-1754) were followed by ^{1}H-^{15}N HSQC spectroscopy, at 700 MHz, at 298 K. One-bond ^{1}H-^{15}N coupling constants were measured by using the IPAP method. RDCs values were calculated as the difference of the fitted $^{1}J_{NH}$ between Ca_4- and Ca_3Ln-CaM N60D.

VI.3 Results and Discussion

Different constructs of six selected CaM targets were cloned with Gateway technology in five expression plasmids carrying different fusion tags: HRPAP20(1-175), CAMTA1 (1547-1618), MAX (22-103), AKAP79 (1-103), (1-72) and (1-170), TUBERIN (1531-1758), RAB3A (18-186). Expression clones were transformed in BL21(DE3)gold, BL21(DE3) Codon Plus RIPL, Origami pLysS and BL21(DE3)C41 pRosetta E. coli strains for expression tests in LB growth medium. The recombinant proteins overexpression was induced with 0.1 and 0.5 mM IPTG and carried out at 37°C and 18°C. Soluble and unsoluble fractions were collected after 3, 6 and 18 hours of induction and all the samples were analysed by SDS-PAGE. Eight constructs shown in Table 1 were selected for scale up and purification tests.

Table 1. Expressed CaM targets.

CaM target	**Fusion Tag**	**Expression** (Soluble-Insoluble)
CAMTA1 (1547-1618)	GST	S
MAX (22-103)	GST	S
	No tag	I
AKAP79 (1-103)	GST	S
TUBERIN (1531-1758)	No tag	I
	His	S
HRPAP20 (1-175)	His	I

HRPAP20 was expressed with the highest yield (Fig. 1), but it was insoluble in any tested conditions. Different refolding protocols were tried, but the protein was not refolded. The GST tagged CAMTA1 and AKAP79 were not cleaved by TEV protease to remove the fusion tag and had a very low yield of expression, while purified GST tagged MAX precipitates when it reaches the concentration of 25 μM, maybe due to protein unfolding.NMR sample of tuberin (1531-1754) was obtained, even if partially degraded (Fig. 2).

Different ways were tried to avoid such problem in tuberin constructs: very fast sample preparation, different protease inhibitors, different purification conditions, cold-inducible vectors, but without results.

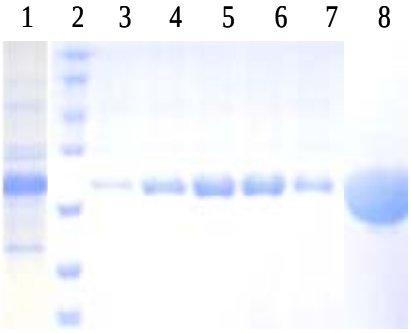


Fig. 1. His tagged HRPAP20 (1-175) expression and purification. Inclusion bodies extract after 6 h expression at 30°C in LB medium, and eluted by Ni^{2+}-loaded HiTrap Chelating column (lane 1), Size exclusion chromatography fractions in denaturing conditions (lanes 3-7), concentrated unfolded sample (lane 8).

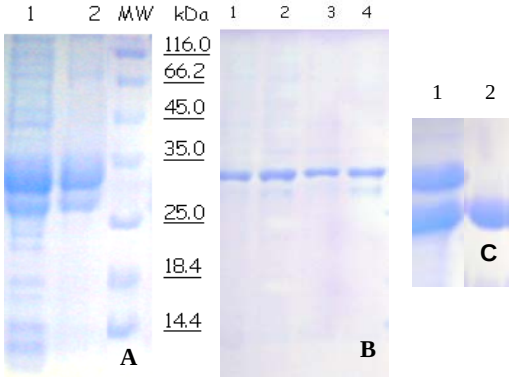


Fig. 2. His tagged Tuberin (1531-1758) expression and purification. ***A:*** soluble extract after 16 h expression at 18°C in LB medium (lane1), elution from His Tag Affinity Column (lane 2).***B:*** Size exclusion chromatography fractions (lanes 1-4). ***C.*** Tuberin degradation, 3 days (lane 1) and 1 month (lane 2) after purification.

Since tuberin loose the C-terminal CaM binding domain because of protease degradation, the (1531-1758) construct was cloned in coexpression vectors with CaM, fused with different fusion tags, in order to test its stability in presence of its partner protein that may protect tuberin from proteases.

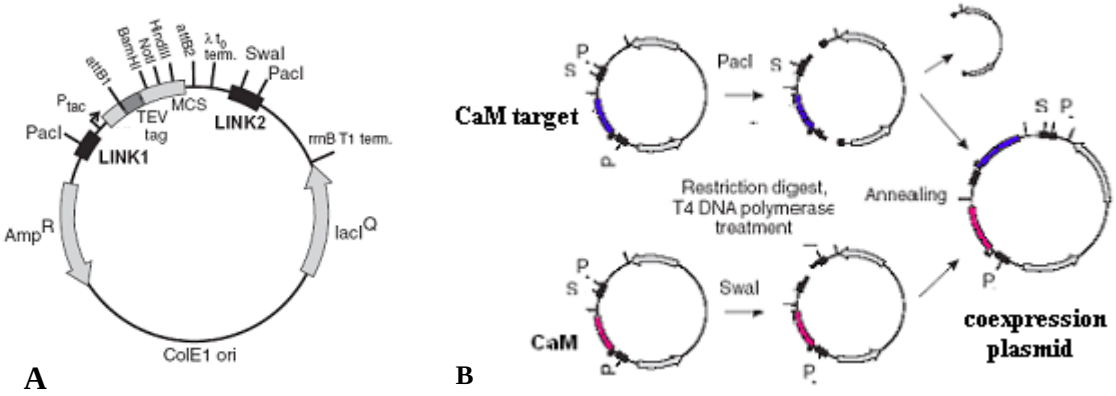


Fig. 3. Map of the pQLink vectors and construction of co-expression plasmids. (A) Map and genetic elements of all pQLink vectors. MCS: multiple cloning site, TEV: TEV protease cleavage site, term: transcription terminator, LINK1 and 2: recombination sites. (B) Construction of a co-expression plasmid from two pQLink plasmids with CaM and CaM target (HRPAP20 or Tuberin) cDNA inserts, respectively.

Also HRPAP20 (1-175) was cloned in coexpression plasmids with CaM, in order to help HRPAP20 folding in presence of soluble CaM. Both tuberin and HRPAP20 were cloned in pQlinks plasmids as described by C. Scheich et al[22], in this way each protein is transcribed by its own promoter, and not in a polycistronic mRNA, that lead to lower expression levels[23]. The cloning protocol is shown in Fig. 3. From the expression tests both tuberin and HRPAP20 were insoluble and in some condition also CaM was expressed in inclusion bodies with the partner protein. Then the His-tagged tuberin construct cloned in pDEST17A Gateway vector was used for further experiments.

In order to have a lower amount of degraded protein, His tagged tuberin (1531-1758) was used for NMR experiments immediately after purifications, indeed during the digestion with TEV protease to remove the tag, the protein degradation was higher. The degradation of tuberin was monitored up to 1 month: after three days, the 50% of the proteins is degraded and in one month the protein is completely degraded (Fig.2C). Tuberin folding was verified by 1D NMR analysis and unlabeled samples were prepared for NMR experiments to test the interaction with CaM. The interaction with CaM was performed with ^{15}N N60D CaM either in the apo and Ca_4 form. ^{1}H-^{15}N HSQC spectra at 298 K of 0.1 mM ^{15}N Ca_4CaM N60D with and without unlabelled tuberin (1531-1758), were performed in a 700 MHz spectrometer with cryoprobe.

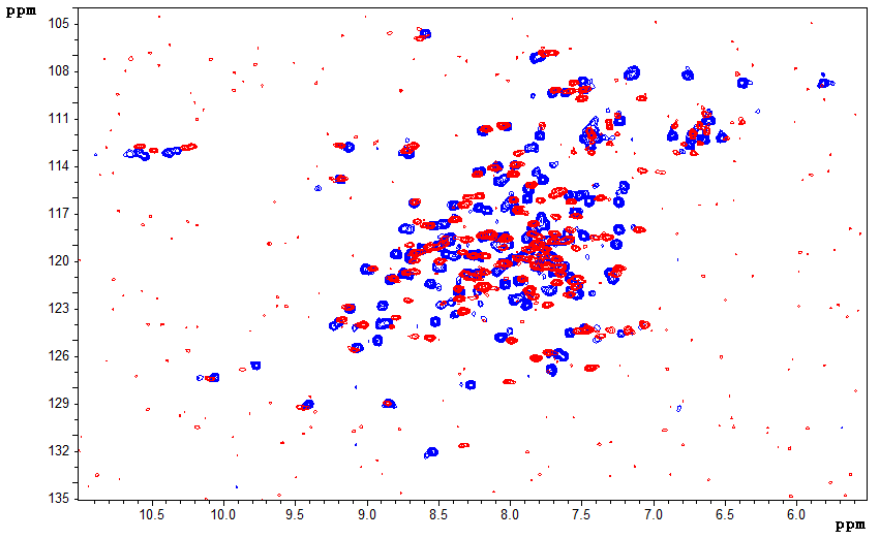


Fig 4. ^{1}H-^{15}N HSQC spectra of Ca4CaM in presence of Tuberin (1531-1758) acqiored at 700MHz spectrometer at 298 K. Samples were prepared in 20mM HEPES pH 7.4, 0.2 M KCl, 1 mM TCEP with 10% $^{2}H_2O$. In blue: ^{1}H-^{15}N HSQC spectrum of 0.1 mM ^{15}N Ca_4CaM without Tuberin (1531-1758). In red: TROSY-^{1}H-^{15}N HSQC spectrum of 0.1 mM ^{15}N Ca_4CaM with unlabelled Tuberin (1531-1758) at 1:4 molar ratio

The titration was performed in five steps, up to 1:4 molar ratio between CaM and tuberin respectively, calculated without considering tuberin degradation. Fig. 4 shows the ^{1}H-^{15}N HSQC spectrum at 298 K of 0.1 mM ^{15}N Ca_4CaM N60D without tuberin (1531-1758) and the TROSY-^{1}H-^{15}N HSQC spectrum of 0.1 mM ^{15}N Ca_4CaM N60D with unlabeled tuberin (1531-1758) at 1:4 molar ratio.

All the samples were in 20 mM HEPES pH 7.4, 0.2 M KCl, 1 mM TCEP. TCEP was used to prevent tuberin aggregation induced by the formation of intermolecular disulfide bridges; TCEP is a reducing agent, more strong and stable than the widely used DTT. During titration, some signals of ^{15}N CaM shifted, and other were not detected, then TROSY-^{1}H-^{15}N HSQC spectra were acquired in order to obtain more signals in the spectra of CaM in complex with tuberin (Fig.4). The data show that CaM and tuberin interact, the exchange between the free and tuberin-bound CaM is fast, due to the interaction between the two proteins in the micromolar range.

The CaM-tuberin complex was analysed in solution by NMR, with a paramagnetic lanthanide substituting the second calcium ion in the N-terminal domain of CaM. Lanthanide ions are used as probes to partially align the CaM molecules in the magnetic field, in order to calculate residual dipolar couplings (RDCs), and evaluate the orientation of the two proteins in the complex. Three different lanthanides were used to prepare Ca_3LnCaM N60D (Ln= Tm, Tb, Yb). Ca_3LnCaM samples were prepared with a titration of 25 mM $CaCl_2$ in the solution of apo CaM, followed by ^{1}H-^{15}N HSQC spectroscopy, to obtain the Ca_3 CaM form, then $LnCl_3$ solutions were added to obtain the Ca_3Ln CaM forms.

The addiction of tuberin (1531-1758) to Ca_3LnCaM induces shifting and widening of some signals of Ca_3LnCaM. At the same time the addition of tuberin induce lost of paramagnetic signals in the spectrum of CaM. This can be due to the binding of Ln^{3+} ions to two Glutamate residues present at the C-terminal of tuberin. To asses this hypothesis, a stop codon was inserted by site-directed mutagenesis after the codon of Isoleucine 1754, so removing the two glutamates from the construct. This new tuberin construct was expressed and prepared with the same protocol. Also the interaction of Ca_3Ln CaM N60D with tuberin was performed in the same conditions. In this case the Ln ions remain bound to CaM. IPAP experiments were performed on 0.2 mM ^{15}N N60D CaM in the Ca_4 and Ca_3Ln^{2+} forms (Ln = Tb, Yb, Tm) in presence of unlabelled 0.4 mM tuberin (1531-1754). RDCs were calculated as the difference of the fitted $^{1}J_{NH}$ between Ca_3LnCaM and Ca_4CaM, in presence of tuberin (1531-1754).

Unfortunately, the large size of the complex that decrease the quality of acquired spectra, prevent a precise measurement of the RDC values. In order to obtain structural information on

the CaM-tuberin complex, a model peptide, corresponding to tuberin (1740-1754), was used to obtain structural information. First it was verified that both tuberin (1531-1754) and (1740-1754) induce the same shifts in ^{15}N Ca_4CaM ^{1}H-^{15}N HSQC spectra.

The binding of Ca_4CaM to the tuberin peptide appears to be slightly stronger than that to tuberin (1531-1754). The estimated binding constant for Ca_4CaM-peptide adduct was of the order of 20 μM. The binding of the tuberin peptide with apoCaM was also detected by NMR spectroscopy, through titration experiments, and a binding constant of about 450 μM has obtained. In fig. 5 the ^{1}H-^{15}N HSQC spectra of CaM, in the Apo- (fig. 5A) and Ca_4-form, in presence of tuberin peptide, are shown (Fig. 5B).

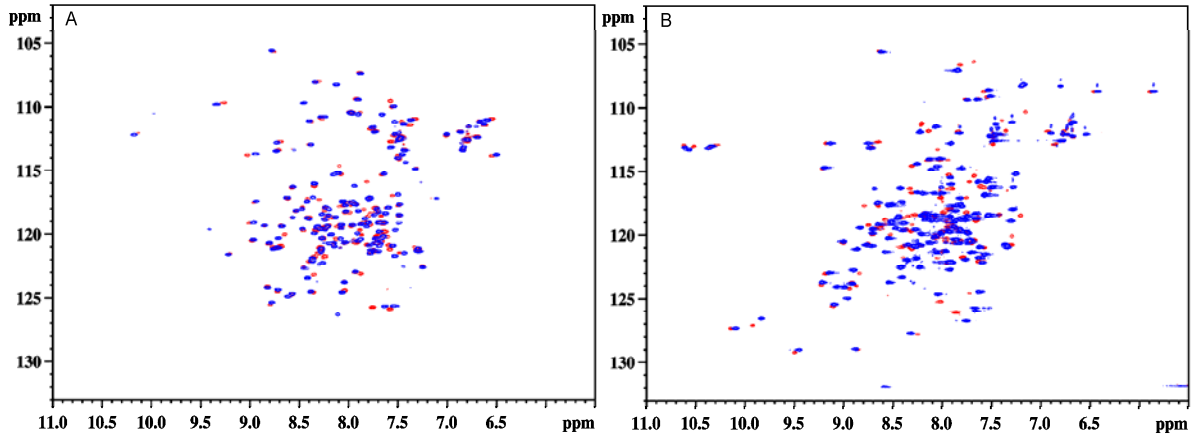


Fig. 5. ^{1}H-^{15}N HSQC spectra of CaM in presence of tuberin (1740-1754) sintetic peptide. (A) ^{1}H-^{15}N HSQC spectrum of 0.45 mM ^{15}N ApoCaM alone (in blue) and in presence of 1.6 mM unlabeled tuberin (1740-1754) (in red). (B) ^{1}H-^{15}N HSQC spectrum of 0.45 mM ^{15}N Ca_4CaM alone (in blue) and in presence of 1.1 mM unlabeled tuberin (1740-1754) (in red)

The higher solubility of the tuberin peptide with respect to the protein allowed the acquisition of NMR spectra on more concentrated samples. ^{1}H-^{15}N HSQC (Fig. 6) and IPAP experiments were performed on 0.45 mM ^{15}N Ca_3Ln CaM in presence of 1.1 mM tuberin peptide (Ln=Tb, Yb, Tm, Dy). The lower molecular weight of the CaM-tuberin peptide complex with respect to the CaM-tuberin (1531-1754) complex, provided a better spectra resolution, a RDC values could be obtained. The analysis of RDCs is still in progress. The experimental RDC values has been used to obtain the magnetic susceptibility tensor for each paramagnetic metal ion, assuming the structural conformation of the N- and C- terminal domains of CaM as previously calculated for other CaM-peptide adducts. The magnetic susceptibility anisotropy $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ were thus obtained for the CaM N-terminal and C-terminal domains separately, for the Tb and Tm ions, as shown in table 2.

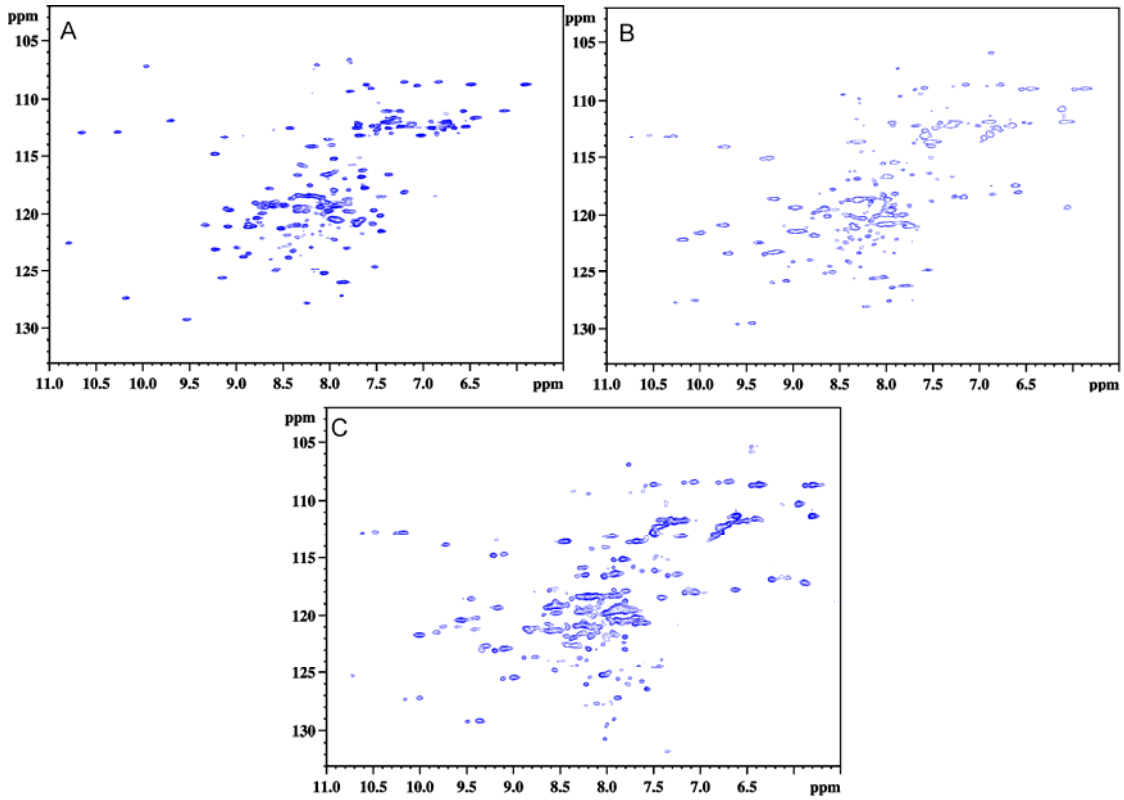


Fig. 6. ^{1}H-^{15}N HSQC spectra of Ca_3LnCaM in presence of tuberin (1740-1754) sintetic peptide. A: $Ca_3YbLnCaM$; B: Ca_3TbCaM; C: Ca_3DyCaM.

		$\Delta\chi_{ax}$ (m^3)	$\Delta\chi_{rh}$ (m^3)
Ca_3TbCaM	N-terminal domain	$4.4 \cdot 10^{-31}$	$1.2 \cdot 10^{-31}$
	C-terminal domain (residues from helices)	$1.9 \cdot 10^{-31}$	$3.9 \cdot 10^{-32}$
Ca_3TmCaM	N-terminal domain	$2.2 \cdot 10^{-31}$	$3.9 \cdot 10^{-31}$
	C-terminal domain (residues from helices)	$1.4 \cdot 10^{-31}$	$7.1 \cdot 10^{-31}$

Table 2. $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ values obtained for Ca_3TbCaM and Ca_3TmCaM in presence of tuberin peptide.

The axial anisotropy values calculated for the Ca_3LnCaM N-terminal were similar to the calues previously obtained for other EF-hand proteins in presence of the same lanthanides. The $\Delta\chi_{ax}$ values obtained from the C-terminal helices residues, were, on the contrary, around 2-3 times smaller, indicating some degree of mobility of the C-terminal domain, with respect to the N-terminal domain of CaM in complex with the tuberin peptide.

Reference List

1. Zhang MJ, Yuan T. Molecular mechanisms of calmodulin's functional versatility. Biochemistry and Cell Biology-Biochimie et Biologie Cellulaire 1998; 76(2-3):313-323.

2. Rhoads AR, Friedberg F. Sequence motifs for calmodulin recognition. FASEB J 1997; 11(5):331-340.

3. Carr DW, Stofkohahn RE, Fraser IDC, Cone RD, Scott JD. Localization of the Camp-Dependent Protein-Kinase to the Postsynaptic Densities by A-Kinase Anchoring Proteins - Characterization of Akap-79. Journal of Biological Chemistry 1992; 267(24):16816-16823.

4. Osawa M, Tokumitsu H, Swindells MB, Kurihara H, Orita M, Shibanuma T et al. A novel target recognition revealed by calmodulin in complex with Ca2+-calmodulin-dependent kinase kinase. Nature Structural Biology 1999; 6(9):819-824.

5. Larsson G, Schleucher J, Onions J, Hermann S, Grundstrom T, Wijmenga SS. A novel target recognition revealed by calmodulin in complex with the basic helix-loop-helix transcription factor SEF2-1/E2-2. Protein Sci 2001; 10(1):169-186.

6. Hermann S, Saarikettu J, Onions J, Hughes K, Grundstrom T. Calcium regulation of basic helix-loop-helix transcription factors. Cell Calcium 1998; 23(2-3):135-142.

7. Onions J, Hermann S, Grundstrom T. A novel type of calmodulin interaction in the inhibition of basic helix-loop-helix transcription factors. Biochemistry 2000; 39(15):4366-4374.

8. Karp CM, Pan HQ, Zhang MY, Buckley DJ, Schuler LA, Buckley AR. Identification of HRPAP20: A novel phosphoprotein that enhances growth and survival in hormone-responsive tumor cells. Cancer Research 2004; 64(3):1016-1025.

9. Karp CM, Shukla MN, Buckley DJ, Buckley AR. HRPAP20: a novel calmodulin-binding protein that increases breast cancer cell invasion. Oncogene 2007; 26(12):1780-1788.

10. Smolarek TA, Wessner LL, McCormack FX, Mylet JC, Menon AG, Henske EP. Evidence that lymphangiomyomatosis is caused by TSC2 mutations: Chromosome 16p13 loss of heterozygosity in angiomyolipomas and lymph nodes from women with lymphangiomyomatosis. American Journal of Human Genetics 1998; 62(4):810-815.

11. Jones AC, Shyamsundar MM, Thomas MW, Maynard J, Idziaszczyk S, Tomkins S et al. Comprehensive mutation analysis of TSC1 and TSC2 - and phenotypic correlations in 150 families with tuberous sclerosis. American Journal of Human Genetics 1999; 64(5):1305-1315.

12. York B, Lou DY, Panettieri RA, Krymskaya VP, Vanaman TC, Noonan DJ. Cross-talk between tuberin, calmodulin, and estrogen signaling pathways. FASEB J 2005; 19(6):1202-+.

13. Daumke O, Weyand M, Chakrabarti PP, Vetter IR, Wittinghofer A. The GTPase-activating protein Rap1GAP uses a catalytic asparagine. Nature 2004; 429(6988):197-201.

14. Inoki K, Li Y, Xu T, Guan KL. Rheb GTPase is a direct target of TSC2 GAP activity and regulates mTOR signaling. Genes & Development 2003; 17(15):1829-1834.

15. Tee AR, Manning BD, Roux PP, Cantley LC, Blenis J. Tuberous sclerosis complex gene products, tuberin and hamartin, control mTOR signaling by acting as a GTPase-activating protein complex toward Rheb. Current Biology 2003; 13(15):1259-1268.

16. Li Y, Corradetti MN, Inoki K, Guan KL. TSC2: filling the GAP in the mTOR signaling pathway. Trends in Biochemical Sciences 2004; 29(1):32-38.

17. Li Y, Inoki K, Guan KL. Biochemical and functional characterizations of small GTPase Rheb and TSC2 GAP activity. Molecular and Cellular Biology 2004; 24(18):7965-7975.

18. York B, Lou DY, Noonan DJ. Tuberin nuclear localization can be regulated by phosphorylation of its carboxyl terminus. Molecular Cancer Research 2006; 4(11):885-897.

19. Noonan DJ, Lou DY, Griffith N, Vanaman TC. A calmodulin binding site in the tuberous sclerosis 2 gene product is essential for regulation of transcription events and is altered by mutations linked to tuberous sclerosis and lymphangioleiomyomatosis. Archives of Biochemistry and Biophysics 2002; 398(1):132-140.

20. Elshorst B, Hennig M, Forsterling H, Diener A, Maurer M, Schulte P et al. NMR solution structure of a complex of calmodulin with a binding peptide of the Ca2+ pump. Biochemistry 1999; 38(38):12320-12332.

21. Bertini I, Gelis I, Katsaros N, Luchinat C, Provenzani A. Tuning the Affinity for Lanthanides of Calcium Binding Proteins. Biochemistry 2003; 42:8011-8021.

22. Scheich C, Kummel D, Soumailakakis D, Heinemann U, Bussow K. Vectors for co-expression of an unrestricted number of proteins. Nucl Acids Res 2007; 35(6).

23. Kim KJ, Kim HE, Lee KH, Han W, Yi MJ, Jeong J et al. Two-promoter vector is highly efficient for overproduction of protein complexes. Protein Sci 2004; 13(6):1698-1703.

VII. SUMMARY AND PERSPECTIVES

NMR spectroscopy and biophysical characterizations techniques, such as native gel electrophoresis, mass spectrometry, light scattering, have been used to characterise protein-protein complexes and mechanisms responsible for protein function.

The *in vitro* characterization of proteins and protein complexes is at the basis of structural studies, since, for these types of experiments, the proteins are not in physiological conditions and are required at very high concentrations, hence, the best experimental conditions have to be found, according to the biophysical characteristics of the proteins.

The *in vivo* study of recombinant protein function helps in the understanding of the physiologic role of a protein, which can be involved in different pathways and exert different functions on the basis of a physiological or pathological context.

Therefore, it is clear that an integrated approach in the study of proteins function is essential, in order to obtain information on proteins and complexes from different techniques, that can provide different point of view on the function of proteins.

Publications

- **Substrate Specificities of Matrix Metalloproteinase 1 in PAR-1 Exodomain Proteolysis. Nesi A, Fragai M. ChemBioChem, (2007) 8:1367-1369**

- **Proteolytic anti-inflammatory activity of MMP13 in liver acute inflammation in animal model of liver fibrosis *(in preparation)***