

***University of Florence***

***International Doctorate in Structural Biology***

***Cycle XXII (2007-2009)***



**Structure, dynamics  
and interaction studies of  
extracellular proteins**

**Ph.D. thesis of**

Maxime Melikian

***Tutor***

***Prof. Ivano Bertini***

***Coordinator***

***Prof. Claudio Luchinat***

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

S.S.D. CHIM/03

« Science sans conscience n'est que ruine de l'âme »

François Rabelais

*This thesis is dedicated to my parents,*

# Table of contents

## Part I : Structure, dynamics and interaction studies of Matrix metalloproteinases

1. Introduction .....	5
1.1 MMPs Architecture .....	6
1.2 Structure and biological activity of the extracellular matrix .....	9
1.3 Pathologies related to MMPs .....	10
1.4 Development of inhibitors .....	12
1.5 Interaction studies between MMPs and ECM components .....	14
1.5.1 Identification of the cleavage site in the triple helical collagen and development of peptide models .....	14
1.5.2 Insights collagenolytic activity : .....	16
1.5.3 Gelatinases follow the degradation process of collagen : .....	18
1.5.4 Matrix metalloproteinases involved in elastin degradation .....	19
1.6 Aim of the research .....	21
1.7 Reference list .....	22
2. Materials and Methods .....	30
2.1 NMR .....	31
2.1.1 Protein assignment experiments .....	31
2.1.2 Residual dipolar coupling .....	31
2.1.3 Pseudo-contact shift .....	32
2.1.4 Relaxation experiments .....	32
2.1.5 NMR titration experiments .....	33
2.2 Assignment strategy .....	34
2.3 Mobility studies from relaxation data .....	35
2.4 Structure calculations .....	36
2.5 Docking Calculation .....	37
2.6. Reference List .....	38
3. Results .....	41
3.1 Evidence of Reciprocal Reorientation of the Catalytic and Hemopexin-Like Domains of Full-Length MMP-12 .....	42
3.2 Interdomain Flexibility in Full-length Matrix Metalloproteinase-1 (MMP-1) .....	54
3.3 Characterization of the MMP-12-elastin adduct .....	62

3.4	Interaction between MMP1 and a collagen peptide model .....	66
3.4.1	<i>Interaction between FL-MMP1 and triple helical peptide (THP) .....</i>	66
3.4.2	<i>Interaction between HPX-MMP1 and THP .....</i>	67
3.4.3	<i>Docking calculation for hemopexin-like domain of MMP1 and THP from chemical shift perturbation using Haddock program .....</i>	68
3.4.4	<i>Docking calculation on catalytic domain of MMP1 and THP (Fields) using Haddock program .....</i>	70
3.4.5	<i>Fitting of MMP1 single domains to the THP from calculated complexes .....</i>	71
3.4.6	<i>FL-MMP1 Structure from THP-fixed single domains (HPX and CAT) using Modeller program .....</i>	72
3.4.7	<i>Docking calculation by Haddock program using chemical shift perturbation on Modeller structure of FL-MMP1 and THP (Fields) .....</i>	73
4.	Conclusion.....	74
 <b>Part II : Study of interaction between S100P and RAGE receptor</b> 		
1.	Introduction .....	79
1.1	Pathologies related to S100P and RAGE activity .....	81
1.2	Structural studies .....	82
1.3	Aim of the project .....	84
2.	Materials and Methods .....	85
2.1	NMR.....	86
2.2	Assignment of the S100P and RAGE V-domain .....	88
2.3	Mobility studies from relaxation data .....	88
2.4	Docking Calculation.....	89
3.	Results .....	90
3.1	NMR backbone assignment of S100P-Ca <sup>2+</sup> form.....	91
3.2	Interaction between S100P protein and V-domain of RAGE receptor .....	93
3.3	Docking calculation on a S100P-RAGE V-domain complex .....	95
3.4	Effects of complex formation on dynamics .....	99
4.	Conclusion.....	101
4.1	Reference list.....	103
	<i>Acknowledgments.....</i>	<i>106</i>
	<i>Curriculum Vitae.....</i>	<i>107</i>

# Part I : Structure, dynamics and interaction studies of matrix metalloproteinases

## 1. Introduction

Matrix metalloproteinases (MMPs) are a class of proteolytic zinc enzymes involved in the degradation of several extracellular proteins, included extracellular matrix components.<sup>1-5</sup> The latter are a heterogeneous group of proteins and glycoproteins forming a structured network that provides a scaffolding for cell organization and mechanical strength.<sup>6-8</sup> Besides the mechanical support to tissues, the extracellular matrix (ECM) regulates the cellular activity, directly or by interplaying with cytokines, growth-factors and receptors.<sup>9,10</sup> ECM exerts these biological activities thanks to the tight binding to the cell cytoskeleton, and to the capability to bind, store and deliver cytokines and growth factors.<sup>11</sup> The degradation of ECM by MMPs has a relevant physiological function for morphogenesis, tissue remodeling and repairing as well as for wound healing.<sup>3,12</sup>

MMPs also process several other extracellular substrates such as extracellular domains of membrane receptors, cytokines and growth factors, as well as other proteases.<sup>13-17</sup> However, the biological effects and the physiological function of this large proteolytic activity is still poorly understood<sup>18</sup>. Recently, intracellular localization has been also reported for MMP1,<sup>19</sup> 2,<sup>20-22</sup> and 11,<sup>23</sup> and a role in proteolysis of intracellular substrates suggested. Because of the potential destructive effects on tissue and cell function, the activity of MMP in healthy tissues is regulated by a tight control of the expression, activation of zymogens precursor,<sup>24</sup> clearance,<sup>25</sup> and by the interaction with physiological inhibitors.<sup>14,26-30</sup>

The uncontrolled or the pathology-driven activity of MMP is associated to a large set of diseases such as rheumatoid arthritis, autoimmune diseases, cancer, etc.<sup>31</sup> The pathological role of MMPs has been largely investigated, and hundreds of inhibitors have been synthesized and tested.<sup>32-36</sup> The research carried out in the last twenty years on MMP inhibitors (MMPI) was affected by a poor knowledge of the biological<sup>37,38</sup> and pathological functions of these multi-domain enzymes and by the high structural similarity among the members of the family that strongly affect the selectivity of the molecules.<sup>39</sup> The progress on molecular studies of MMPs function and structure are now changing the perspectives for a real application of MMPI in the treatment of human diseases.<sup>40</sup>

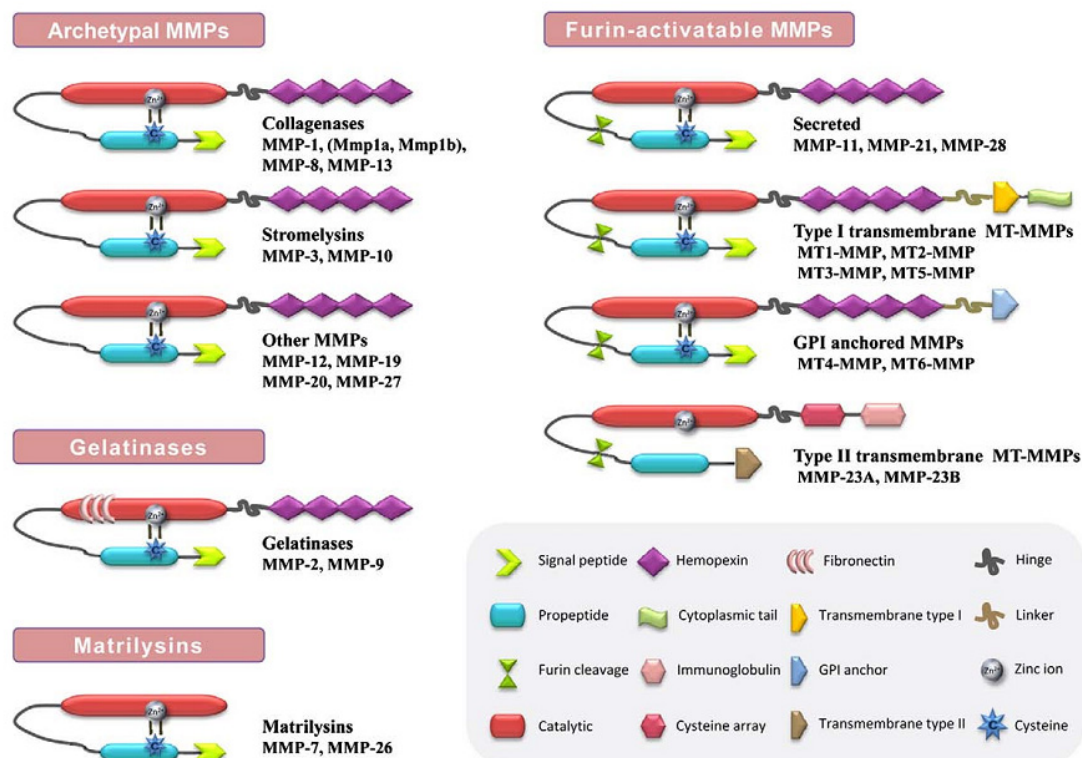
## 1.1 MMPs Architecture

Matrix metalloproteinases also called matrixins belong to the category of Zinc-endopeptidases included in the metzincin superfamily. To date, 24 members of MMP family have been identified. These enzymes contain a three histidines motif which coordinates a zinc metal ion inside the catalytic domain (CAT). A glutamic acid located close to the zinc ion is also required for catalytic activity. An additional methionine completes the following pattern HExxHxxxxxHxxxxxxxM identified as a signature for metalloproteinases.

MMP classification can be done in several ways. For example, a common classification consists in grouping the MMPs according to their respective substrates such as collagen, elastin, fibronectin...

The present figure classifies the MMPs according to their domain organisation. MMPs can be regrouped into four groups :

- \_ Archetypal MMPs
- \_ Matrilysins
- \_ Gelatinases
- \_ Furin activatable MMPs



**Figure 1.** The mammalian family of matrix metalloproteinases. Structural classification of MMPs based on their domain arrangement. <sup>41</sup>

### *Archetypal MMPs :*

This large group can be divided into three sub-groups corresponding to the respective substrates : collagenases composed by MMP1, MMP8 and MMP13, stromelysins composed by MMP3 and MMP10 and other enzymes which are MMP12, MMP19, MMP20 and MMP27.

Regarding collagenases, it was reported these proteases are not active in absence of their hemopexin-like (HPX) domain.<sup>42</sup> Then collagenolytic activity required the intervention of both the HPX domain and the CAT domain.

Stromelysins share a similar domain architecture respect with collagenases, have broad substrate specificity but are unable to perform cleavage on triple helical collagen. However, they have the property to activate collagenases by removing the propeptide<sup>43</sup>. In addition, stromelysins are able to release active form MMP9 by performing cleavage on the prodomain.<sup>44</sup>

Among other archetypal MMPs, macrophage metallo-elastase (MMP12) mainly located in the macrophage is able to degrade a broad range of ECM substrates. MMP12 has the best ability of degrading elastin but is also able to process several other molecules like aggrecan, fibronectin, laminin and type IV collagen.

### *Gelatinases :*

One of the main structural characteristics of gelatinases MMP2 and MMP9 consist in the presence of an additional fibronectin domain composed by three sub-units inserted in the CAT domain. These sub-units play an important role in the degradation of gelatine which corresponds to a denaturated form of collagen.<sup>45</sup> In addition gelatinases may degrade many other different components from ECM such as elastin, fibronectin, laminin and aggrecan.<sup>4</sup> Else, HPX domain of MMP9 is involved in the mechanism of cell migration.<sup>46</sup> Gelatinolysis activity is important in angiogenesis through the intervention of different activating factors.<sup>47</sup>

### *Matrilysins :*

Matrilysin subgroup is composed by only two protein members MMP7 and MMP26. These proteins are constituted by a single CAT domain. Matrilysin targets are ECM components in particular type IV collagen, laminin, entactin and elastin.

### *Furin activatable MMPs :*

These proteins contain a common specific sequence between the propeptide and the catalytic domain which is recognized by convertase proteases allowing further activation of these MMPs. This category of MMP can be subdivided into three subgroups : secreted MMPs, membrane type MMPs and type II transmembrane MMP. The first subgroup is composed by MMP11, MMP21 and MMP28. MMP11 is quite similar to stromelysins in terms of architecture and function and participates in the degradation of collagen VI. In addition, the protein is playing also a role in wound healing. MMP21 is active during embryo development of kidney, intestine and skin. However, this protein hasn't been extensively studied to date. In the same way, MMP28 was also identified in several cancers but the target remains unknown.<sup>48</sup>

Many of the furin activatable MMPs belong to the category of the membrane type MMPs and are also called MT1-MMP, MT2-MMP, MT3-MMP, MT4-MMP, MT5-MMP and MT6-MMP. These proteins are characterized by a transmembrane domain. Different patterns distinguish these MT-MMPs. The first group composed by MT1-MMP, MT2-MMP, MT3-MMP and MT5-MMP contains a hydrophobic transmembrane region followed by a cytoplasmic tail which has been suggested to be involved activation mechanism. Other membrane type MMPs (MT4-MMP and MT6-MMP) contain a GPI (glycosyl phosphatidyl inositol) anchor motif in the transmembrane region. Due to the sub-cellular localisation of these MT-MMPs at the surface of the cell, these proteins play a specific role of regulating the cell environment. In particular, MT-MMPs are important in cell attachment. Like other MMP members, MT-MMPs are able to process different ECM components such as gelatine, collagen IV and fibronectin. Interestingly, MT-MMPs are both expressed in normal and tumour cells. Depending on the expression level and the cellular type ; they may induce or reduce tumour progression.<sup>41</sup>

Regarding MMP23A and MMP23B, the proteins have the same amino acid sequences but are encoded by two distinct genes in the human genome. Differences can be found in the gene promoter which controls the protein expression. Then, the same protein can be found at a different concentration in different tissues. From the topology, MMP23 A and B are the unique members lacking a peptide signal, a cysteine switch motif and the hemopexin domain. In the other hand, they contain a cysteine array and an immunoglobulin domain. Inversely respect with the MT-MMPs, the transmembrane domain is located at the N-terminal part of the protein. Regarding the role of MMP23 A and B, it can be noticed they are expressed in ovary, testis and prostate.<sup>49</sup>

This suggests a probable role in reproduction process; however, the protein activity in vivo still has to be investigated.

### *Evolutionary events*

The MMPs domain organisation is the consequence of several evolutionary events. Indeed, MMPs are also present in invertebrates and plants. From the CAT domain sequences, it appears MMPs from invertebrates are more similar to plants than to vertebrates, suggesting a very ancient origin of this family protein.<sup>50</sup> Moreover, HPX domain is not present in plants.

To date, the main theory regarding MMP family evolution suggests the existence of a common gene ancestor coding for the CAT domain which had duplicated and evolved in order to generate the actual MMP family. Then, the multi-domain organisation of the protein would be originally the consequence of a fusion of two genes. Interestingly, association between a CAT domain and another structural domain has increased the protein specificity.

Matrilysins MMP7 and MMP26 appeared from a deletion of HPX domain. This evolutionary event constitutes an exception respect with the main tendency of MMPs to recruit additional structural domains.<sup>51</sup>

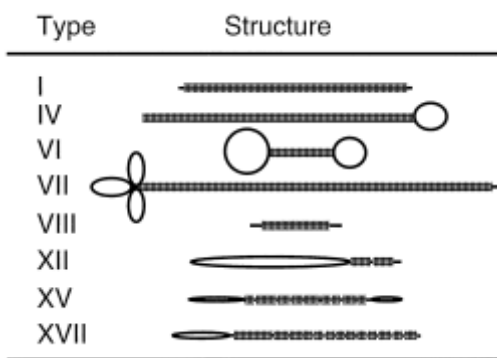
The most conserved regions in MMPs correspond to the CAT domain. In the other hand, less conserved sequences may correspond to important protein regions for specificity in particular the HPX domain which is possibly involved in the substrate recognition.



## 1.2 Structure and biological activity of the extracellular matrix

Extracellular matrix (ECM) is constituted by a complex of biomolecules assembled in a network. Inside this biological architecture, each molecule is related to a specific function. Interestingly, as in many others biological systems the function is highly correlated to the structure. The major components of the ECM are collagen molecules, glycoproteins and proteoglycans. Association of these different components forms heterochimeric complexes.

Collagen molecules are the most abundant molecules of the ECM. They consist mainly in repeated G-X-Y motifs in different chains which assemble to form triple helical structures. The following figure represents the different types of collagen.



**Figure 2.** Schematic representation of typical collagen molecules.<sup>6</sup>

Thick dotted lines : helical rods;  
Circles, ellipses : NC domains.

Only the processed molecules are shown for collagen I and VII

Due to its triple helical structure, collagen forms a very robust and stable base for the matrix allowing then the binding of additional components. The biomolecular complexes formed become resistant enough to accommodate the cell organisation forming the different tissues. In particular, collagen is the most abundant proteins in bones (more than 90%). Some collagen fragments were identified in fossils by mass spectrometry. In order to illustrate the collagen robustness, it can be mentioned that some collagen fragments were extracted from bones of a 68-million-year-old *Tyrannosaurus Rex*.<sup>52</sup>

Some particular motifs such as fibronectin or von Willebrand factor<sup>53</sup> can be found inside the different chains forming the collagen. ECM also contains non collagenous molecules in particular some glycoproteins such as elastin or fibronectin.<sup>6</sup>

Fibronectin (Fn) is composed by a repetition of three different motifs (Fn I, Fn II and Fn III) which are connected together in a single amino-acid chain. Several subunit repetition structures have been resolved both by X-ray crystallography and magnetic resonance.<sup>54</sup> Interestingly, three fibronectin type II modules are present in gelatinases MMP2 and MMP9. This structure similarity found between an ECM component and gelatinases should be important for interaction.

Elastin, as its name indicates shows some elasticity properties and represents around 2-4 % of skin composition.<sup>55</sup> This protein contains in particular a lot of cross-links and repetition of hydrophobic residues VPGVG which renders the protein particularly insoluble.

The fibrils which form the ECM are composed by a heterogeneous assemblage of different types of triple helical collagen. These fibrils, in addition to collagenous fragments are connected also to proteoglycans. Then, proteoglycans are able to connect collagen fibrils

to the elastin network. Other connections were observed between fibronectin domains and collagen fibrils mediated by interactions with proteoglycans.<sup>56</sup>

ECM formation can be initiated by a self-polymerisation of type IV collagen and followed by a self-assembly mechanism of laminin 1. These steps may provide a scaffold to allow the binding of other members which form the extracellular matrix.

Once the ECM formation is complete, cell binding function to the matrix is enabled. Interaction between ECM components and cell receptors regulates several cellular activities : in particular, cell adhesion, migration, differentiation, proliferation, and apoptosis. Cell adhesion function was observed in some fibronectin modules but also for some collagen fragments which have been mapped.<sup>57</sup>

In order to emphasize the ECM importance, a loss of cell adhesion may lead to several diseases, in particular cancer development due to cell proliferation.

### **1.3 Pathologies related to MMPs**

Matrix metalloproteinases are involved in several processes such as cell proliferation, cell migration, cell adhesion, embryogenesis, angiogenesis, wound healing and bone development. MMPs activity is tightly regulated in several ways :

- \_ A gene can be activated by hormones, cytokines and growth factor.
- \_ At the DNA level, histone acetylation or DNA methylation may occur.
- \_ At the RNA level, mRNA stability may be a factor of modulation.
- \_ At the protein level, a pro-domain maintain the protein in an inactive form.<sup>41</sup>
- \_ Active protein may be inhibited naturally by the intervention of TIMPs (tissue inhibitor of matrix metalloproteinases).<sup>58</sup>

In vivo experiments based on gain and loss of functions have demonstrated the role of MMP in the development of diseases.

#### *Cancer :*

MMPs are believed to promote tumour progression by initiating carcinogenesis, enhancing tumour angiogenesis, disrupting local tissue architecture to allow tumour growth and finally by breaking down basement membrane barriers for metastatic spread.

Initial studies emphasized the degradation of type IV collagen, a major component of basement membranes by MMP-2 and MMP-9 in tumour tissues. Subsequently, other findings demonstrated ECM components could be processed by different MMPs rendering the interpretation more complex. Most MMPs in tumours are produced by stromal cells rather than cancer cells. In fact, MMP expression by fibroblasts can be stimulated by a cell surface glycoprotein expressed by tumour cell.<sup>59</sup>

#### *Inflammatory diseases :*

Involvement of MMP1, MMP3 and MMP9 were observed in rheumatoid and osteoarthritis. At the same time, an important role for aggrecanase, a member of the ADAM family has been proposed in articular damage.

### *Cardiovascular disease :*

There has been a long standing interest in the role of MMPs in cardiovascular disease. Numerous studies have demonstrated increased levels of MMPs at sites of atherosclerosis and aneurism formation.<sup>60</sup> The idea that inflammatory process could be connected to a development of atherosclerosis plaque was suggested by extracellular matrix degradation induced by macrophages. In this mechanism, macrophages are known to secrete and activate MMPs provoking then inflammation.

### *Lung disease*

Elevated MMPs levels have been implicated in various lung pathologies, including acute respiratory distress syndrome, asthma, bronchiectasis and cystic fibrosis. In particular, MMPs and TIMPs are produced in many cells from lung tissues rendering more complicated the analysis of individual metalloproteinase in lung disease.

### *Central nervous system disease*

From observations obtained on animal models resembling multiple sclerosis and Guillain-Barre's syndrome, MMP9 role has been evidenced in different types of neurologic diseases.<sup>61</sup> Some treatment using MMP inhibitors has reversed the pathology in animal models of brain injury.

### *Shock syndromes*

MMP8 and MMP9 are located in the granules of polymorphonuclear leukocytes. These cells are key effectors in inflammatory and infectious processes. A role for these MMPs in shock is supported by studies in MMP9 deficient mice shown to be resistant to endotoxic shock.<sup>62</sup>

### *Bone disorders*

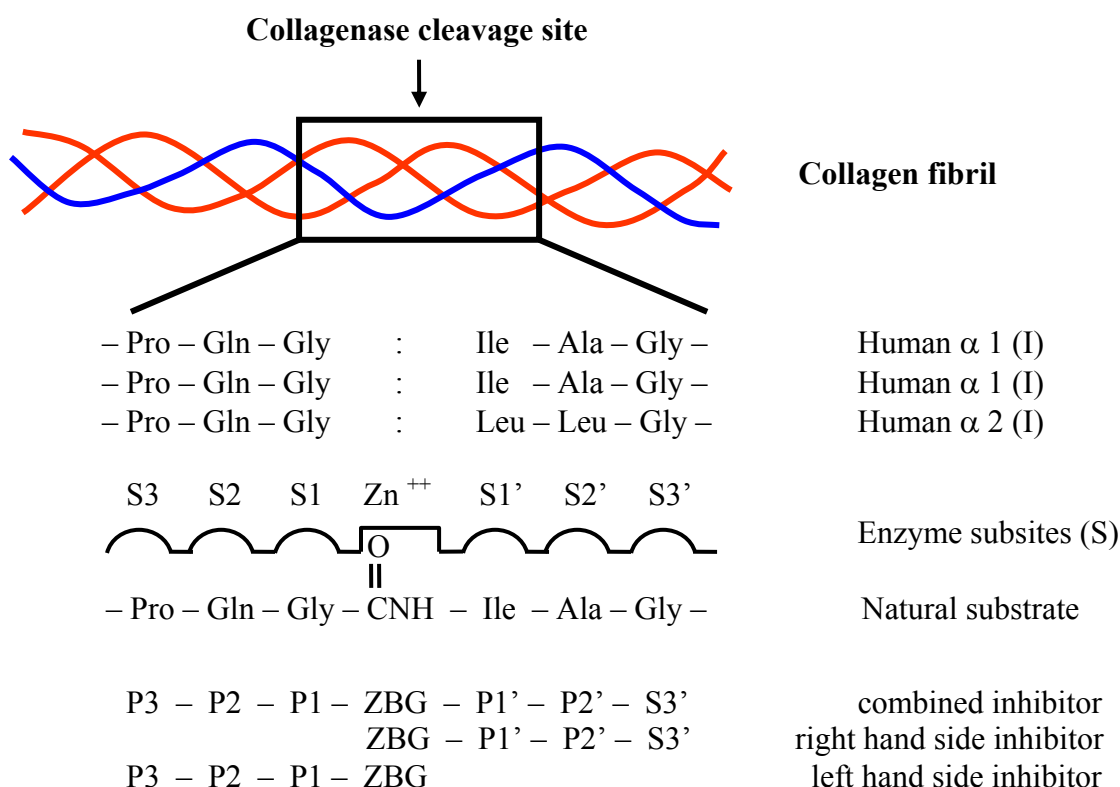
Mutations in MMPs may cause several disorders in the protease activity leading to an altered ECM turnover in function. For example, in MMP2, a Y244X mutation causes an abrupt end of transcription due to the presence of a codon stop in RNA and subsequently the absence of the following functional domains.<sup>63</sup> This non functional protein results in osteolysis and bone disorders.

### *Single nucleotide Polymorphism (SNP) modifying MMP expression*

In rare cases, some disease can appear due to dramatic changes in MMP expression caused by a single nucleotide polymorphism (SNP) in DNA. The protein remains functional but regulation of protein expression is affected. For example, insertion of an additional Guanine in the MMP1 gene promoter increases gene expression by providing an additional binding site for a transcription factor. The consequence is an increase of mRNA transcription leading to a high MMP1 expression in particular in colorectal and lung cancer.<sup>64</sup>

## 1.4 Development of inhibitors

The over-expression of one or more MMPs in several pathological tissues has prompted the development of synthetic inhibitors since the early 80'. In particular, high-level expression of various MMPs were observed in several cancer cell lines and in the stromal cells of the tissue surrounding the tumor mass.<sup>65,66</sup> These experimental evidences and the idea that the breaking of the extracellular matrix components was crucial to enable tumor cells to invade the neighboring tissue and to reach blood vessels<sup>67,68</sup> have suggested MMPs as potential targets for cancer therapy. The active site on the CAT domain is an obvious target to inhibit the enzymatic activity, and was first considered to develop new MMP inhibitors<sup>69</sup> Succinate peptidomimetic hydroxamates based on substrate sequences were developed at the end of the 80' and then tested in clinical trials against several cancer types.<sup>70-72</sup> For collagenases, the cleavage site sequence of collagen was used to design the corresponding inhibitors (figure 3). This cleavage site was determined by mass spectrometry and Edman degradation from the released fragments after collagenolysis.



**Figure 3.** Design of matrix metalloproteinase inhibitors based on the collagen cleavage site.<sup>32</sup> ZBG : zinc binding group.

These molecules bind the active site in a way that resembles the substrate in the transition state, establishing a network of hydrogen bonds with the protein backbone, fitting the deep S1' pocket with a lipophilic moiety and binding the catalytic zinc ion in a bidentate fashion.<sup>73,74</sup> Although these hydroxamates are potent inhibitors of the catalytic domain, they have broad-spectrum inhibitory activity and modest stability in vivo.<sup>32</sup> Sulfonamide hydroxamate derivatives constitute the largest class of MMPs inhibitors.<sup>32,36,75,76</sup> The general formula contains a hydroxamic group designed to bind the catalytic zinc ion, and a large hydrophobic moiety targeted to the S<sub>1</sub>' cavity. The two groups are tethered by a sulfonamide moiety forming hydrogen bonds with the binding site.<sup>77,78</sup> Their easy synthetic accessibility explains the popularity of these molecules among medicinal chemists. Unfortunately, they suffer from most of the same drawbacks of the peptidomimetic hydroxamates.<sup>39</sup>

In particular, the hydroxamic moiety has been long considered the zinc binding group (ZBG) of choice for MMP inhibitors (MMPI) thanks to its unique capability to bind the catalytic metal ion in the active site of matrix metalloproteinases.<sup>79</sup> However, hydroxamate is metabolically labile<sup>80</sup> and it contributes to increase toxicity due to its low metal binding selectivity.<sup>81</sup> In the last two decades several non-hydroxamate inhibitors have been developed.<sup>36</sup> The carboxylic acid has been often selected as ZBG in MMPI, despite its relatively weak affinity for the zinc ion.<sup>82</sup> The use of a carboxylic acid to replace the hydroxamic acid warrants higher stability of the inhibitor under physiological conditions,<sup>80</sup> improves the solubility and provides better bioavailability.<sup>83</sup> Following the auto-inhibitory activity of prodomain, several thiol-based ligands have been also designed and tested.<sup>84-86</sup> As the hydroxamate derivatives, also the compounds with a thiol such as ZBG are broad-spectrum MMPI but with a better oral availability. Pyrimidinetrione derivatives<sup>87,88</sup> constitute a further class of orally-available MMPI and, though rather unselective, they seem to cause fewer adverse neuromuscular effects. Phosphinic and phosphonate moieties, mimicking the gem-diol group of the transition state, are good ZBGs.<sup>79,89,90</sup> Some phosphorus-based ligands exhibit a good selectivity towards specific matrix metalloproteinases but the problems related to a generally poor oral availability still have to be solved.<sup>91</sup> Hydroxypyrrone, hydroxypyridinone and hydroxy-urea are further examples of ZBG exploited to design MMPI.<sup>92,93</sup> In particular, the structural features of hydroxypyrrone-, hydroxypyridinone-based ZBGs seems to provide a relevant contribution to the observed selectivity in this class of inhibitors.<sup>81</sup> It has been hypothesized that several side-effects observed during the clinical trials with MMPI could be related to the poor specificity and selectivity of the investigated molecules.<sup>39</sup> These include musculoskeletal pain and inflammation which often required the interruption of the therapy. Several attempts to design selective inhibitors have been carried out by using the structure-based strategy.<sup>94,95-97</sup> Although this strategy has been successfully applied in drug discovery against several protein targets,<sup>98,99</sup> in case of MMPs the results have been relatively disappointing. The poor results of the structure-based approach in designing of MMPI can be likely due not only to the high homology of the MMP's members but also to the poor knowledge of the dynamical features of these proteins.<sup>100</sup>

Actually, the structural data considered in structure-based drug design resulting from X-ray structures doesn't take into account protein flexibility. The detrimental effect of protein flexibility for MMPI selectivity is particularly evident for MMP1, where the small and shallow S1' pocket can accommodate a large hydrophobic moiety by reorienting arginine 214,<sup>101</sup> but it has been also observed in MMP-12 where internal conformational adjustments occur upon binding of the inhibitor.<sup>102</sup>

X-ray crystallography can provide high resolution protein structures, although it cannot sample the conformational heterogeneity related to protein mobility but only the structure corresponding to the energy minimum in those experimental conditions. Complementary information can be achieved by NMR structures, where the poor definition of specific protein regions together with dynamic information provide evidence of internal flexibility.<sup>103-108</sup> Recently, the active sites of several MMPs have been structurally characterized by X-ray and NMR and in all investigated proteins exhibits highly similar structural and dynamical features. A comparative analysis carried out on solution and crystal structures of different MMP shows that several regions of the catalytic domain undergo relatively wide and collective motions.<sup>109</sup> The analysis of the RMSD per residue performed on NMR structures of different MMPs shows that all the investigated proteins are disordered in the same loop regions.

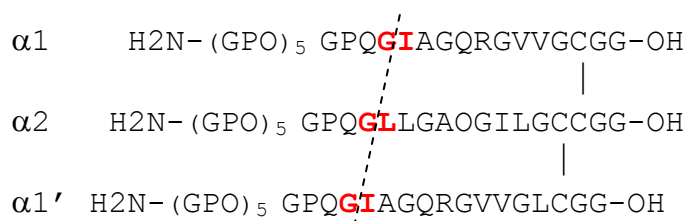
## 1.5 Interaction studies between MMPs and ECM components

### 1.5.1 Identification of the cleavage site in the triple helical collagen and development of peptide models

In parallel with development of MMP inhibitors, a deeper investigation of the interaction between MMPs and their respective substrates has been carried out. Cleavage sites composed by a Gly-Ile-Ala and Gly-Leu-Ala motif were identified in the triple helical region of type II and type III collagen. A similar cleavage site was found also in type I collagen.<sup>110</sup>

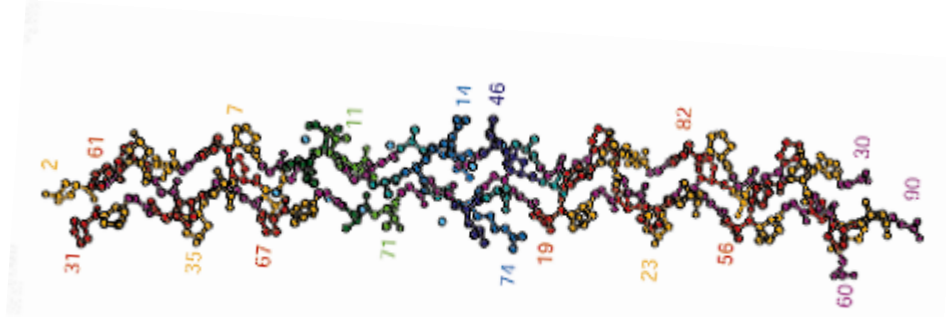
Further studies<sup>111</sup> showed collagen can be cleaved into  $\frac{3}{4}$  and  $\frac{1}{4}$  fragments specifically by collagenases. Moreover, it appears that triple helical structure of collagen cannot be cleaved by other broad-specificity proteases like trypsin, elastase, or thermolysin but only by collagenases. Indeed, to ensure the ECM stability, collagen molecule shouldn't be sensitive to degradation by a non specific peptidase. These results underlined the ability of collagenases to degrade collagen contained into fibrils despite the low solubility of collagen.

In order to study interaction between proteins and collagen fibrils, triple helical peptide (THP) collagen models were synthesised. Peptide assembly was achieved by building disulfide bridges in order to bring the three chains together into a heterotrimer (figure 4).<sup>112</sup>



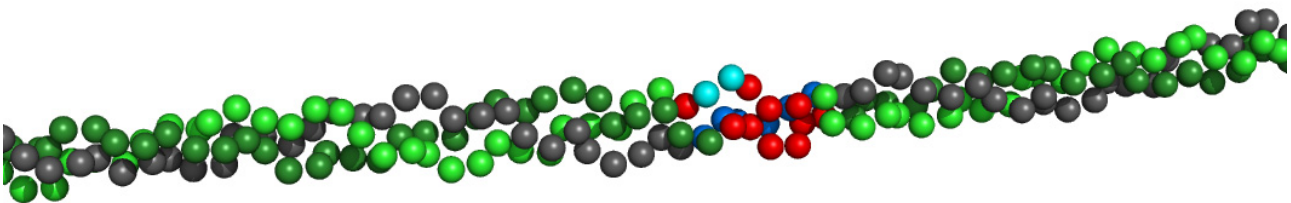
**Figure 4.** Sequences of a heterotrimeric collagenous peptide containing the collagenase cleavage site P4-P'<sub>9/10</sub> of the  $\alpha 1$ ,  $\alpha 2$  and  $\alpha 1'$  chain of type I collagen flanked at the C-terminus by an artificial cystine- knot and at the N-terminus by Gly-Pro-Hyp repeats known to induce and stabilize the triple helical conformation.<sup>112</sup>

A crystal structure of a collagen like type III model peptide was solved<sup>113</sup> (figure 5) in which triple helical structure is maintained along the entire fragment containing the cleavage site. This structure may contradict a previous hypothesis which suggested a local unfolding around the cleavage site allowing collagenases to perform cleavage on collagen. However, the region around cleavage site may not be completely rigid in solution.



**Figure 5.** A partial sequence from human type III collagen and the sequence of the T3-785 peptide. The sequence (Ile 785–Gly 796) included in the peptide is highlighted and shown in red.<sup>113</sup>

Considering the structure above, enzymatic cleavage may not occur without a local unwinding around the cleavage site by collagenases. One mechanism proposed was the triple helical collagen unwinding would be obtained by removal of water molecules. The consequence would be an opening of the triple helix by destabilizing hydrogen bonds which maintain the structure. However, such mechanism is uncertain because it would lead to multiple cleavage sites whereas only one cleavage site was found on type I collagen after hydrolysis by MMP1 collagenase.



**Figure 6.** Zoom on the cleavage site of Type I collagen microfibrillar structure from pdb 3HQV.<sup>114</sup>

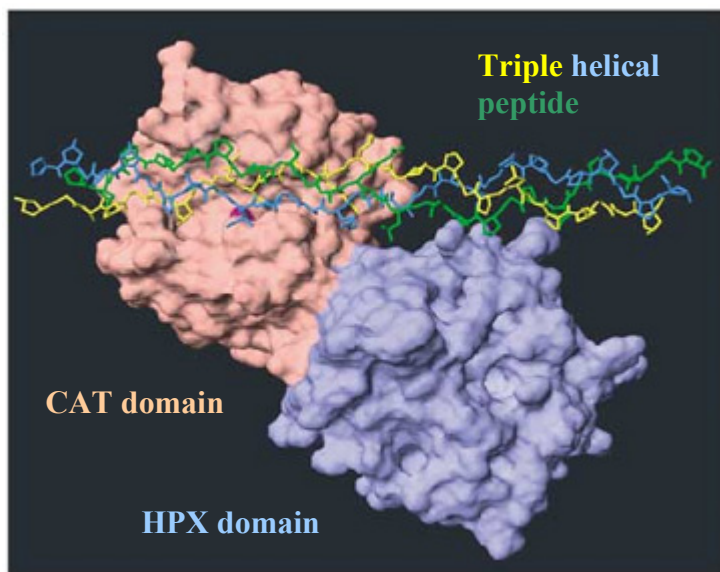
A low resolution structure of a microfibrillar type 1 collagen *in situ* has been solved (figure 6). This 3D structure confirms a triple helical organisation for collagen in the fibril unit.

### 1.5.2 Insights collagenolytic activity :

Regarding catalytic mechanism, unwinding process is necessary before cleavage considering the collagenases active site in CAT domains is too narrow to accommodate an entire triple helix collagen (figure 7). About unwinding mechanism, a cooperative intervention of CAT and HPX domain was suggested considering each domain is required. Still this mechanism has not been completely elucidated.

Moreover, the interaction model described between MMP1 and collagen peptide<sup>115</sup> based on X-ray structures doesn't take into account the possible motions between, CAT, HPX domain and linker peptide in MMP1 and motions inside the triple helical collagen. Actually, NMR studies showed triple helical peptides showed some backbone mobility close to the cleavage site at a temperature of 25°C.<sup>116</sup>

Interestingly, at a physiological temperature of 37°C, MMP1 is able to process efficiently collagen but at a temperature lower than 25°C, cleavage of gelatine corresponding to denaturated collagen is more efficient than cleavage of collagen which confirms activity of collagenases depends both on the temperature and on substrate behaviour.

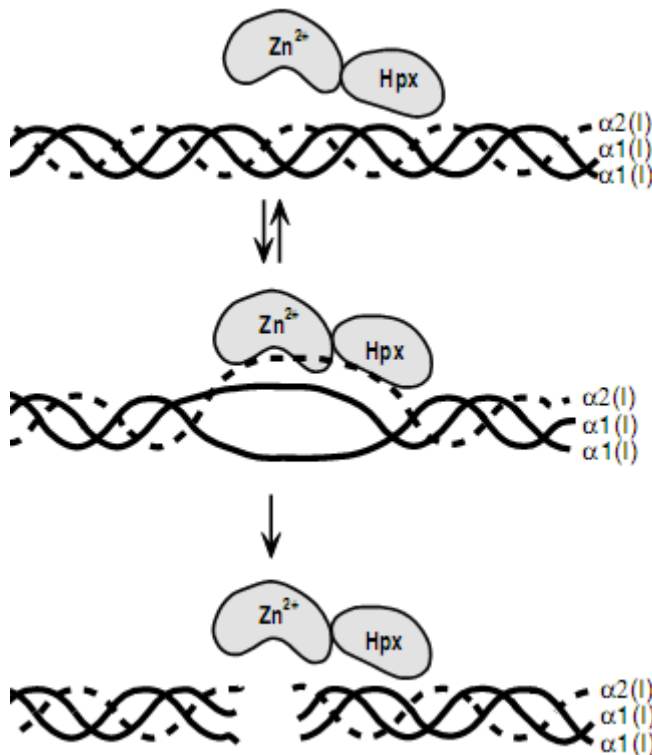


**Figure 7.** Alignment of the triple-helical peptide with the active site of MMP1.<sup>117</sup> Collagen triple-helical peptides<sup>113</sup> were manually aligned into the active site of the catalytic domain of porcine MMP1<sup>118</sup> using Insight II/Discover and the image was produced with Swiss PDB viewer.<sup>119</sup>

Finally, the following mechanism has been proposed.<sup>117</sup> The description consisted into two distinguished steps where the first is to unwind of the triple helical collagen and the second is the cleavage by the collagenases. Experiments were performed using a E219A MMP1 mutated in the active site. Mutant alone was unable to cut collagen but cleavage activity was restored by supplying the system with another wild type CAT domain which is normally unable to perform the cleavage alone. Another interesting result obtained by denaturation experiment in presence of MMP1 showed that collagen unwinding occurs only locally. This result is consistent with the fact that only one cleavage site was identified for MMP1 on type 1 collagen.



Interestingly, collagenase activity can be obtained by the reassociation of isolated CAT and HPX domain using a higher concentration of the isolated domains respect with the full-length (FL) protein. This result indicates both CAT and HPX domains are essential and complementary for collagenolysis. According to these data, linker peptide function consists in enhancing the cooperation between the two structural domains by increasing entropic contribution.



**Figure 8.** Steps involved in collagenolysis. Collagenase binds to and locally unwinds collagen before it cleaves the triple-helical interstitial collagen.<sup>117</sup>

To summarise this collagenolysis mechanism, MMP1 first binds the triple helical collagen, unwinds locally around the cleavage site and hydrolyses subsequently the three chains which composes the collagen (figure 8).

This mechanism is also probably facilitated by a reduced density of hydroxyprolines around the Gly-Ile cleavage site.<sup>120</sup> By looking at the residue sequence of type III collagen for example for both chains  $\alpha1$  and  $\alpha2$ , it appears that dipeptide Gly-Ile can be found several times. Nevertheless, there is a unique cleavage site for MMP1. Then, unrolling capability of the triple helical chain could depend on a local low density of hydroxyprolines. Mutations around the cleavage site replacing natural amino-acids by hydroxyprolines render the protein resistant to hydrolysis. However, it was shown that collagen binding by MMP1 was retained.

In addition to collagen binding, HPX domain is responsible for other functions. From the pro-MMP1 crystal structure,<sup>121</sup> it is evidenced the prodomain interacts also with the HPX domain.

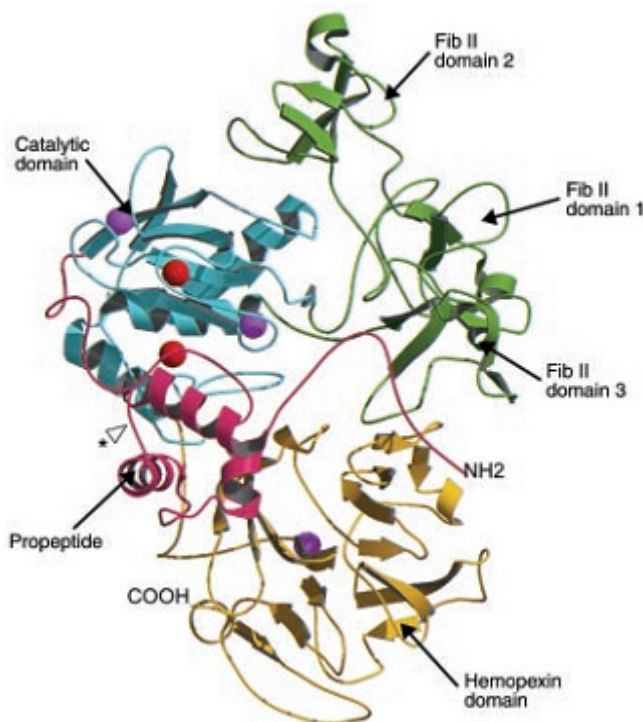
For MMP2, HPX domain is involved in a cascade reaction leading to the enzyme activation.<sup>122</sup>

### 1.5.3 Gelatinases follow the degradation process of collagen :

After collagen hydrolysis by collagenases, gelatinases like MMP2 or MMP9 can continue the degradation on unfolded collagen also called gelatine. MMP2 cleavage activity during blood coagulation of fibrinogen glycoprotein is mediated by its HPX domain. Removal of this domain doesn't abolish the hydrolysis but reduces dramatically the catalytic activity.<sup>123</sup> Furthermore, hydrolysis of fibrinogen in the absence of the HPX domain gives rise to additional cleavage sites. This indicates HPX domain, may function as a guide to address MMPs to their respective targets.

Some experiments have shown HPX-MMP9 can also bind gelatine.<sup>124</sup> Isolated HPX domain in solution inhibits the enzyme activity by competition with the full length MMP9.

In addition to CAT and HPX domain, MMP2 and MMP9 contain three sub-unit of fibronectin-like II domain. These three sub-units have a high similarity with the fibronectin ECM component able to bind collagen.<sup>56</sup> Interestingly, fibronectin binding site to collagen corresponds also to the cleavage site by collagenases around G775-I776 from the type I collagen sequence. Moreover, mutations in this area affect fibronectin binding but also fibrillogenesis.



**Figure 9.** Structure of proMMP2. The pro-domain, catalytic domain, fibronectin domains, and hemopexin domain are shown in red, blue, green, and yellow, respectively.  $Zn^{2+}$  ions are indicated in red, and  $Ca^{2+}$  ions are magenta. Asterisk indicates the cleavage site for MT1-MMP.<sup>125</sup>

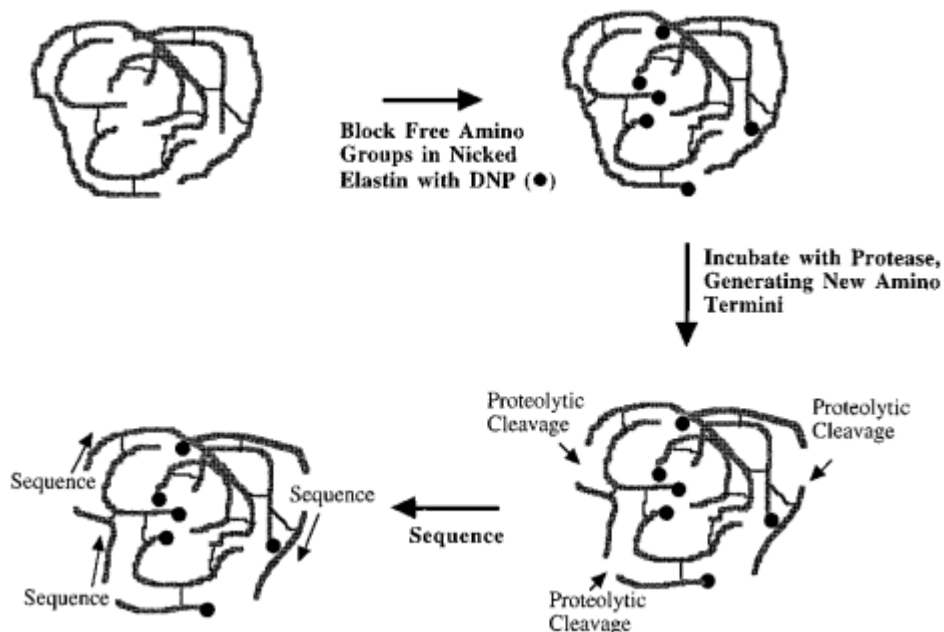
The three fibronectin II like modules are able to bind denatured collagen type I (gelatine I), IV and V and elastin.<sup>126</sup> The role of the interaction between fibronectin-like modules and gelatine type I was investigated.<sup>127</sup> Recombined enzyme MMP2 lacking these three modules has a significantly reduced efficiency respect with the full-length (FL) protein. In addition, experiments performed in presence of both FL-MMP2 and fibronectin like modules alone resulted in competition for interaction with substrate.

#### 1.5.4 Matrix metalloproteinases involved in elastin degradation

Elastin is an important component of ECM, its function is to provide elasticity to tissues in particular in the skin. From its nature of a cross-linked compound with hydrophobic properties, elastin is very stable and highly resistant to most of the non-specific proteases.

Among the MMP family, gelatinases MMP2 and MMP9, matrilysins and macrophage metalloelastase MMP12 are responsible for elastolytic activity<sup>128</sup>. For MMP12, HPX domain is not indispensable for this activity.<sup>129</sup> For matrilysins, the enzyme is only composed by the CAT domain. For gelatinases, it was suggested only CAT domain is responsible for elastin degradation. However, experiments performed using constructs containing only the CAT domain showed MMP2 and MMP9 without their fibronectin II like modules are unable to degrade elastin. Activity was restored then by using constructs containing both CAT domain with insertion of fibronectin-like modules. Nevertheless, for elastin degradation by gelatinases HPX domain doesn't seem to be crucial.

Comparing all energy requirements for collagenolysis, gelatine and elastin hydrolysis by their respective enzymes has revealed elastin is the substrate which has the lowest activation barrier<sup>130</sup> with respect to other enzymatic cleavages. Probably, energy level depends on the nature of the substrate. In particular, more conformational changes are necessary for hydrolysis of a triple helix in collagen rather than in elastin were cleavage sites are directly accessible. Enzymatic assay in presence of elastin and several enzymes was performed. Elastin was successfully degraded by human MMP12 human, murine MMP12, MMP9 and a matrilysin but not by collagenase MMP1 despite a high similarity with MMP12.



**Figure 10.** Strategy for characterizing proteinase cleavage sites in insoluble elastin.<sup>130</sup>

By using dinitrophenol (DNP) labelling, it was possible to identify several amino acids located at the different cleavage sites. Interestingly, the most abundant residues found were Gly, Leu and Ile which are also present in the cleavage site of collagen. In addition, some Val and Ala residues were also identified.

MMP12 is one of the most active enzymes involved in elastin degradation. Respect with MMP1 which recognises a unique cleavage site on a particular triple helical structure of collagen, or with MMP2 which may recognise two or three cleavage sites,<sup>117</sup> MMP12 recognises a larger diversity of substrates from ECM such as elastin, fibronectin, laminin, entacin and also type IV collagen.<sup>129</sup> This broad spectrum of substrates catalysed by MMP12 may explain its requirement for macrophage activity where the protein is essentially found.

As already mentioned, MMP12 actively degrades elastin but was found also to activate other MMPs such as MMP2 and MMP3.<sup>131</sup> In a recent work, a detailed characterisation of MMP12 activity towards elastin was performed in order to identify cleaved fragments of elastin<sup>132</sup> to obtain a map of the cleavages sites in the protein sequence. Physiologically, MMP12 is expressed at a very low concentration in the skin. To study elastin *in vitro* a partial hydrolysis by chemical agents is required for solubilisation considering its cross-linked nature and its high hydrophobic content. Fragment peptides from degradation by CAT-MMP12 were collected, separated and analysed by mass spectrometry. Peptides sequences were also confirmed by *de novo* sequencing. Forty one fragments were identified from 4 to 41 residues which allowed locating cleavage site in the human elastin protein sequence. Similarly to the previous results,<sup>130</sup> several Gly-Leu and Gly-Ile cleavage sites were identified. In addition, these sites could be inserted in the context of elastin sequence (Swiss Prot P15502) which allows taking into account the fragment with residues surrounding. Else a better understanding of recognition site by MMP12 becomes available.

MMP12 also degrades tropoelastin which is the precursor of cross-linked elastin.<sup>133</sup> Expression of tropoelastin is enhanced in response to UV damage which occurs on skin.<sup>134</sup> Respect with elastin, tropoelastin is more soluble and also more exposed to cleavage by Cat-MMP12. Then, more cleavage sites were found for tropoelastin respect with hydrolysis of CAT-MMP12 on elastin but surprisingly with longer average fragments which mainly overlap. Presence of more cleavage sites could be explained partially because tropoelastin in not cross-linked, then more sites may be accessible to the enzyme.

## 1.6 Aim of the research




The present research has been carried out to clarify the molecular mechanisms by which MMPs degrade structurally organized ECM components such as collagen and elastin. Actually, the available crystal structures of collagenase MMP1 show a compact arrangement of the two domains, which are connected by a 14 aa linker. Docking calculations performed on these structures suggest that the interaction with triple helical collagen should occur in a way that can lead to first unwinding and then cleavage of individual filaments. In particular, it has been recently suggested that such concerted action could occur much more easily if the two domains could enjoy at least a partial conformational independence.

In this research, NMR experiments have been performed on MMP1 and MMP12 to assign hydrogen, nitrogen and carbon resonances of the proteins and to monitor protein dynamics. The NMR analysis has been integrated with small angle X-ray scattering (SAXS) data and, in case of MMP12, also with crystallographic information.

Then, the research has been focused on the recognition by MMP1 and MMP12 of their natural substrates, collagen and elastin respectively. The interaction between MMP12 and a soluble mixture of elastin fragments has been investigated by NMR and docking calculations. The results have provided the structural determinants relevant for the binding of MMP12 to elastin cleavage sites.

Also, the interaction between MMP1 and a synthetic triple helical collagen peptide has been investigated by NMR during this research. Actually, the proteolytic steps occurring during collagen hydrolysis by MMP1 were not elucidated. In particular, the role of the HPX domain, and the possible role of the interdomain flexibility were debated. The interaction of the selected collagen model with MMP1 monitored by NMR has provided evidence for the role of the interdomain flexibility in collagen unwinding and hydrolysis allowing identifying the protein regions involved in collagen recognition and cleavage.

The present thesis is based on the following articles :

-  Bertini I, Calderone V, Fragai M, Jaiswal R, Luchinat C, Melikian M, Mylonas E, Svergun DI. Evidence of reciprocal reorientation of the catalytic and hemopexin-like domains of full-length MMP-12.  
*J Am Chem Soc.* **130** (22) : 7011-21. (2008).
-  Bertini I, Fragai M, Luchinat C, Melikian M, Mylonas E, Sarti N, Svergun DI. Interdomain flexibility in full-length matrix metalloproteinase-1 (MMP-1).  
*J Biol Chem.* **284** (19) : 12821-8 (2009).
-  Bertini I, Fragai M, Luchinat C, Melikian M, Venturi C. Characterisation of the MMP-12-elastin adduct.  
*Chemistry.* **15** (32):7842-5. (2009).

## 1.7 Reference list

1. Shapiro,S.D. Matrix metalloproteinase degradation of extracellular matrix: biological consequences. *Curr. Opin. Cell Biol.* **10**, 602-608 (1998).
2. Woessner,J.F.Jr. & Nagase,H. Matrix Metalloproteinases. *J. Biol. Chem.* **274**, 21491-21494 (1999).
3. Page-McCaw,A., Ewald,A.J., & Werb,Z. Matrix metalloproteinases and the regulation of tissue remodelling. *Nat. Rev. Mol. Cell Biol.* **8**, 221-233 (2007).
4. Overall,C.M. Molecular determinants of metalloproteinase substrate specificity. *Mol. Biotechnol.* **22**, 51-86 (2002).
5. Visse,R. & Nagase,H. Matrix metalloproteinases and tissue inhibitors of metalloproteinases: structure, function, and biochemistry. *Circ Res* **92**, 827-839 (2003).
6. Aumailley,M. & Gayraud,B. Structure and biological activity of the extracellular matrix. *J. Mol. Med.* **76**, 253-265 (1998).
7. Bosman,F.T. & Stamenkovic,I. Functional structure and composition of the extracellular matrix. *J. Pathol.* **200**, 423-428 (2003).
8. Tanzer,M.L. Current concepts of extracellular matrix. *J. Orthop. Sci.* **11**, 326-331 (2006).
9. Eble,J.A. The extracellular matrix in health and disease. *Curr. Pharm. Des.* **15**, 1275-1276 (2009).
10. Sternlicht,M.D. & Werb,Z. How matrix metalloproteinases regulate cell behavior. *Annual Review of Cell and Developmental Biology* **17**, 463-516 (2001).
11. Berrier,A.L. & Yamada,K.M. Cell-matrix adhesion. *J. Cell Physiol* **213**, 565-573 (2007).
12. Stamenkovic,I. Extracellular matrix remodelling: the role of matrix metalloproteinases. *J. Pathol.* **200**, 448-464 (2003).
13. Boire,A., Covic,L., Agarwal,A., Jacques,S., Sherif,S., & Kuliopulos,A. PAR1 is a matrix metalloprotease-1 receptor that promotes invasion and tumorigenesis of breast cancer cells. *Cell* **120**, 303-313 (2005).
14. Nagase,H., Visse,R., & Murphy,G. Structure and function of matrix metalloproteinases and TIMPs. *Cardiovasc. Res.* **69**, 562-573 (2006).
15. Parks,W.C., Wilson,C.L., & Lopez-Boado,Y.S. Matrix metalloproteinases as modulators of inflammation and innate immunity. *Nat. Rev. Immunol.* **4**, 617-629 (2004).
16. D'Alessio,S., Fibbi,G., Cinelli,M., Guiducci,S., Del Rosso,A., Margheri,F., Serrati,S., Pucci,M., Kahaleh,B., Fan,P.S., Annunziato,F., Cosmi,L., Liotta,F., Matucci-Cerinic,M., & Del Rosso,M. Matrix metalloproteinase 12-dependent cleavage of urokinase receptor in systemic sclerosis microvascular endothelial cells results in impaired angiogenesis. *Arthritis Rheum.* **50**, 3275-3285 (2004).
17. Fragai,M. & Nesi,A. Substrate Specificities of Matrix Metalloproteinase 1 in PAR-1 Exodomain Proteolysis. *ChemBioChem* **8**, 1367-1369 (2007).
18. Doucet,A. & Overall,C.M. Protease proteomics: revealing protease in vivo functions using systems biology approaches. *Mol. Aspects Med.* **29**, 339-358 (2008).
19. Limb,G.A., Matter,K., Murphy,G., Cambrey,A.D., Bishop,P.N., Morris,G.E., & Khaw,P.T. Matrix metalloproteinase-1 associates with intracellular organelles and confers resistance to lamin A/C degradation during apoptosis. *Am. J. Pathol.* **166**, 1555-1563 (2005).

20. Wang, W., Schulze, C.J., Suarez-Pinzon, W.L., Dyck, J.R., Sawicki, G., & Schulz, R. Intracellular action of matrix metalloproteinase-2 accounts for acute myocardial ischemia and reperfusion injury. *Circulation* **106**, 1543-1549 (2002).
21. Sawicki, G., Leon, H., Sawicka, J., Sariahmetoglu, M., Schulze, C.J., Scott, P.G., Szczesna-Cordary, D., & Schulz, R. Degradation of myosin light chain in isolated rat hearts subjected to ischemia-reperfusion injury: a new intracellular target for matrix metalloproteinase-2. *Circulation* **112**, 544-552 (2005).
22. Kwan, J.A., Schulze, C.J., Wang, W., Leon, H., Sariahmetoglu, M., Sung, M., Sawicka, J., Sims, D.E., Sawicki, G., & Schulz, R. 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.* **18**, 690-692 (2004).
23. Luo, D., Mari, B., Stoll, I., & Anglard, P. Alternative splicing and promoter usage generates an intracellular stromelysin 3 isoform directly translated as an active matrix metalloproteinase. *J. Biol. Chem.* **277**, 25527-25536 (2002).
24. Ra, H.J. & Parks, W.C. Control of matrix metalloproteinase catalytic activity. *Matrix Biol.* **26**, 587-596 (2007).
25. Barmina, O.Y., Walling, H.W., Fiocco, G.J., Freije, J.M., Lopez-Otin, C., Jeffrey, J.J., & Partridge, N.C. Collagenase-3 binds to a specific receptor and requires the low density lipoprotein receptor-related protein for internalization. *J. Biol. Chem.* **274**, 30087-30093 (1999).
26. Drouin, L., Overall, C.M., & Sodek, J. Identification of matrix metalloendoproteinase inhibitor (TIMP) in human parotid and submandibular saliva: partial purification and characterization. *J. Periodontal Res.* **23**, 370-377 (1988).
27. Overall, C.M. & Sodek, J. Reciprocal regulation of collagenase, 72 kDa-gelatinase, and TIMP gene expression and protein synthesis in human fibroblasts induced by concanavalin A. *Matrix Suppl* **1**, 209-211 (1992).
28. Strongin, A.Y., Marmor, B.L., Grant, G.A., & Goldberg, G.I. Plasma membrane-dependent activation of the 72-kDa type IV collagenase is prevented by complex formation with TIMP-2. *J. Biol. Chem.* **268**, 14033-14039 (1993).
29. Uria, J.A., Ferrando, A.A., Velasco, G., Freije, J.M., & Lopez-Otin, C. Structure and expression in breast tumors of human TIMP-3, a new member of the metalloproteinase inhibitor family. *Cancer Res.* **54**, 2091-2094 (1994).
30. Arumugam, S., Hemme, C.L., Yoshida, N., Suzuki, K., Nagase, H., Berjanskii, M., Wu, B., & Van Doren, S.R. TIMP-1 Contact Sites and Perturbations of stromelysin 1 Mapped by NMR and a Paramagnetic Surface Probe. *Biochemistry* **37**, 9650-9657 (1998).
31. Wolf, K., Wu, Y.I., Liu, Y., Geiger, J., Tam, E., Overall, C., Stack, M.S., & Friedl, P. Multi-step pericellular proteolysis controls the transition from individual to collective cancer cell invasion. *Nat. Cell Biol.* **9**, 893-904 (2007).
32. Whittaker, M., Floyd, C.D., Brown, P., & Gearing, A.J. Design and Therapeutic Application of Matrix Metalloproteinase Inhibitors. *Chem. Rev.* **99**, 2735-2776 (1999).
33. Skiles, J.W., Gonnella, N.C., & Jeng, A.Y. The design, structure, and clinical update of small molecular weight matrix metalloproteinase inhibitors. *Curr Med Chem* **11**, 2911-2977 (2004).
34. Rao, B.G. Recent developments in the design of specific Matrix Metalloproteinase inhibitors aided by structural and computational studies. *Curr. Pharm. Des* **11**, 295-322 (2005).
35. Overall, C.M. & Kleifeld, O. Towards third generation matrix metalloproteinase inhibitors for cancer therapy. *British Journal of Cancer* **94**, 941-946 (2006).

36. Aureli,L., Gioia,M., Cerbara,I., Monaco,S., Fasciglione,G.F., Marini,S., Ascenzi,P., Topai,A., & Coletta,M. Structural bases for substrate and inhibitor recognition by matrix metalloproteinases. *Curr. Med. Chem.* **15**, 2192-2222 (2008).
37. Vincenti,M.P. & Brinckerhoff,C.E. Signal transduction and cell-type specific regulation of matrix metalloproteinase gene expression: can MMPs be good for you? *J. Cell Physiol* **213**, 355-364 (2007).
38. Martin,M.D. & Matrisian,L.M. The other side of MMPs: Protective roles in tumor progression. *Cancer Metastasis Rev.* **26**, 717-724 (2007).
39. Coussens,L.M., Fingleton,B., & Matrisian,L.M. Matrix Metalloproteinase Inhibitors and Cancer: Trials and Tribulations. *Science* **295**, 2387-2392 (2002).
40. Lopez-Otin,C. & Matrisian,L.M. Emerging roles of proteases in tumour suppression. *Nat. Rev. Cancer* **7**, 800-808 (2007).
41. Fanjul-Fernandez,M., Folgueras,A.R., Cabrera,S., & Lopez-Otin,C. Matrix metalloproteinases: Evolution, gene regulation and functional analysis in mouse models. *Biochim. Biophys. Acta* (2009).
42. Murphy,G. & Knauper,V. Relating matrix metalloproteinase structure to function: why the "hemopexin" domain? *Matrix Biol.* **15**, 511-518 (1997).
43. Barksby,H.E., Milner,J.M., Patterson,A.M., Peake,N.J., Hui,W., Robson,T., Lakey,R., Middleton,J., Cawston,T.E., Richards,C.D., & Rowan,A.D. Matrix metalloproteinase 10 promotion of collagenolysis via procollagenase activation: implications for cartilage degradation in arthritis. *Arthritis Rheum.* **54**, 3244-3253 (2006).
44. Geurts,N., Martens,E., Van,A., I, Proost,P., Opendakker,G., & Van den Steen,P.E. Beta-hematin interaction with the hemopexin domain of gelatinase B/MMP-9 provokes autocatalytic processing of the propeptide, thereby priming activation by MMP-3. *Biochemistry* **47**, 2689-2699 (2008).
45. Gomis-Ruth,F.X. Catalytic domain architecture of metzincin metalloproteases. *J. Biol. Chem.* **284**, 15353-15357 (2009).
46. Dufour,A., Sampson,N.S., Zucker,S., & Cao,J. Role of the hemopexin domain of matrix metalloproteinases in cell migration. *J. Cell Physiol* **217**, 643-651 (2008).
47. Egeblad,M. & Werb,Z. New functions for the matrix metalloproteinases in cancer progression. *Nat. Rev. Cancer* **2**, 161-174 (2002).
48. Werner,S.R., Dotzlaf,J.E., & Smith,R.C. MMP-28 as a regulator of myelination. *BMC. Neurosci.* **9**, 83 (2008).
49. Velasco,G., Pendas,A.M., Fueyo,A., Knauper,V., Murphy,G., & Lopez-Otin,C. Cloning and characterization of human MMP-23, a new matrix metalloproteinase predominantly expressed in reproductive tissues and lacking conserved domains in other family members. *J. Biol. Chem.* **274**, 4570-4576 (1999).
50. Massova,I., Kotra,L.P., Fridman,R., & Mobashery,S. Matrix Metalloproteinases: structures, evolution, and diversification. *FASEB J.* **12**, 1075-1095 (1998).
51. Puente,X.S., Sanchez,L.M., Overall,C.M., & Lopez-Otin,C. Human and mouse proteases: a comparative genomic approach. *Nat. Rev. Genet.* **4**, 544-558 (2003).
52. Asara,J.M., Schweitzer,M.H., Freimark,L.M., Phillips,M., & Cantley,L.C. Protein sequences from mastodon and Tyrannosaurus rex revealed by mass spectrometry. *Science* **316**, 280-285 (2007).
53. Chu,M.L., Zhang,R.Z., Pan,T.C., Stokes,D., Conway,D., Kuo,H.J., Glanville,R., Mayer,U., Mann,K., Deutzmann,R., & . Mosaic structure of globular domains in the human type VI collagen alpha 3 chain:



- similarity to von Willebrand factor, fibronectin, actin, salivary proteins and aprotinin type protease inhibitors. *EMBO J.* **9**, 385-393 (1990).
54. Dickinson,C.D., Veerapandian,B., Dai,X.P., Hamlin,R.C., Xuong,N.H., Ruoslahti,E., & Ely,K.R. Crystal structure of the tenth type III cell adhesion module of human fibronectin. *J. Mol. Biol.* **236**, 1079-1092 (1994).
  55. Rosenbloom,J., Abrams,W.R., & Mecham,R. Extracellular matrix 4: the elastic fiber. *FASEB J.* **7**, 1208-1218 (1993).
  56. Dzamba,B.J., Wu,H., Jaenisch,R., & Peters,D.M. Fibronectin binding site in type I collagen regulates fibronectin fibril formation. *J. Cell Biol.* **121**, 1165-1172 (1993).
  57. Vandenberg,P., Kern,A., Ries,A., Luckenbilledds,L., Mann,K., & Kuhn,K. Characterization of A Type-Iv Collagen Major Cell Binding-Site with Affinity to the Alpha-1-Beta-1 and the Alpha-2-Beta-1 Integrins. *Journal of Cell Biology* **113**, 1475-1483 (1991).
  58. Maskos,K. & Bode,W. Structural basis of matrix metalloproteinases and tissue inhibitors of metalloproteinases. *Mol. Biotechnol.* **25**, 241-266 (2003).
  59. Yan,L., Zucker,S., & Toole,B.P. Roles of the multifunctional glycoprotein, emmprin (basigin; CD147), in tumour progression. *Thromb. Haemost.* **93**, 199-204 (2005).
  60. Van den Steen,P.E., Dubois,B., Nelissen,I., Rudd,P.M., Dwek,R.A., & Opdenakker,G. Biochemistry and molecular biology of gelatinase B or matrix metalloproteinase-9 (MMP-9). *Crit Rev. Biochem. Mol. Biol.* **37**, 375-536 (2002).
  61. Kieseier,B.C., Clements,J.M., Pischel,H.B., Wells,G.M., Miller,K., Gearing,A.J., & Hartung,H.P. Matrix metalloproteinases MMP-9 and MMP-7 are expressed in experimental autoimmune neuritis and the Guillain-Barre syndrome. *Ann. Neurol.* **43**, 427-434 (1998).
  62. Dubois,B., Starckx,S., Pagenstecher,A., Oord,J., Arnold,B., & Opdenakker,G. Gelatinase B deficiency protects against endotoxin shock. *Eur. J. Immunol.* **32**, 2163-2171 (2002).
  63. Martignetti,J.A., Aqeel,A.A., Sewairi,W.A., Boumah,C.E., Kambouris,M., Mayouf,S.A., Sheth,K.V., Eid,W.A., Dowling,O., Harris,J., Glucksman,M.J., Bahabri,S., Meyer,B.F., & Desnick,R.J. Mutation of the matrix metalloproteinase 2 gene (MMP2) causes a multicentric osteolysis and arthritis syndrome. *Nat. Genet.* **28**, 261-265 (2001).
  64. Zhu,Y., Spitz,M.R., Lei,L., Mills,G.B., & Wu,X. A single nucleotide polymorphism in the matrix metalloproteinase-1 promoter enhances lung cancer susceptibility. *Cancer Res.* **61**, 7825-7829 (2001).
  65. Stetler-Stevenson,W.G. & Hewitt,R. Matrix Metalloproteinases and tumor invasion:from correlation and causality to the clinic. *Sem. Cancer Biol.* **7**, 147-154 (1996).
  66. McCawley,L.J. & Matrisian,L.M. Matrix metalloproteinases: multifunctional contributors to tumor progression. *Mol. Med. Today* **6**, 149-156 (2000).
  67. Liotta,L.A., Tryggvason,K., Garbisa,S., Hart,I., Foltz,C.M., & Shafie,S. Metastatic potential correlates with enzymatic degradation of basement membrane collagen. *Nature* **284**, 67-68 (1980).
  68. Ossowski,L. & Reich,E. Antibodies to plasminogen activator inhibit human tumor metastasis. *Cell* **35**, 611-619 (1983).
  69. Moore,W.M. & Spilburg,C.A. Peptide hydroxamic acids inhibit skin collagenase. *Biochem. Biophys. Res. Commun.* **136**, 390-395 (1986).
  70. Brown,P.D. Ongoing trials with matrix metalloproteinase inhibitors. *Expert. Opin. Investig. Drugs* **9**, 2167-2177 (2000).

71. Steward,W.P. & Thomas,A.L. Marimastat: the clinical development of a matrix metalloproteinase inhibitor. *Expert. Opin. Investig. Drugs* **9**, 2913-2922 (2000).
72. Pavlaki,M. & Zucker,S. Matrix metalloproteinase inhibitors (MMPi): the beginning of phase I or the termination of phase III clinical trials. *Cancer Metastasis Rev* **22**, 177-203 (2003).
73. Lang,R., Kocourek,A., Braun,M., Tschesche,H., Huber,R., Bode,W., & Maskos,K. Substrate specificity determinants of human macrophage elastase (MMP-12) based on the 1.1 angstrom crystal structure. *Journal of Molecular Biology* **312**, 731-742 (2001).
74. Grams,F., Reinemer,P., Powers,J.C., Kleine,T., Pieper,M., Tschesche,H., Huber,R., & Bode,W. X-ray structures of human neutrophil collagenase complexed with peptide hydroxamate and peptide thiol inhibitors. Implications for substrate binding and rational drug design. *Eur. J. Biochem.* **228**, 830-841 (1995).
75. Floyd,C.D., Lewis,C.N., Patel,S.R., & Whittaker,M. A method for the synthesis of hydroxamic acids on solid phase. *Tetrahedron Lett.* **37**, 8045-8048 (1996).
76. MacPherson,L.J., Bayburt,E.K., Capparelli,M.P., Carroll,B.J., Goldstein,R., Justice,M.R., Zhu,L., Hu,S., Melton,R.A., Fryer,L., Goldberg,R.L., Doughty,J.R., Spirito,S., Blancuzzi,V., Wilson,D., O'Byrne,E.M., Ganu,V., & Parker,D.T. Discovery of CGS 27023A, a non-peptidic, potent, and orally active stromelysin inhibitor that blocks cartilage degradation in rabbits. *J. Med. Chem.* **40**, 2525-2532 (1997).
77. Moy,F.J., Chanda,P.K., Chen,J.M., Cosmi,S., Edris,W., Skotnicki,J.S., Wilhelm,J., & Powers,R. NMR solution structure of the catalytic fragment of human fibroblast collagenase complexed with a sulfonamide derivative of a hydroxamic acid compound. *Biochemistry* **38**, 7085-7096 (1999).
78. Bertini,I., Calderone,V., Fragai,M., Luchinat,C., Mangani,S., & Terni,B. Crystal structure of the catalytic domain of human matrix metyalloproteinase 10. *J. Mol. Biol.* **336**, 707-716 (2004).
79. Castelhana,A.L., Billedeau,R., Dewdney,N., Donnelly,S., Horne,S., Kurz,L.J., Liak,T.J., Martin,R., Uppington,R., Yuan,Z., & Krantz,A. Novel indolactam-based inhibitors of matrix metalloproteinases. *Bioorg. Med. Chem. Lett.* **5**, 1415-1420 (1995).
80. Wang,X.Q., Choe,Y.C., Craik,C.S., & Ellman,J.A. Design and synthesis of novel inhibitors of gelatinase B. *Bioorg. Med. Chem. Lett.* **12**, 2201-2204 (2002).
81. Agrawal,A., Romero-Perez,D., Jacobsen,J.A., Villarreal,F.J., & Cohen,S.M. Zinc-Binding Groups Modulate Selective Inhibition of MMPs. *ChemMedChem* **3**, 812-820 (2008).
82. Pikul,S., Ohler,N.E., Ciszewski,G., Laufersweiler,M.C., Almstead,N.G., De,B., Natchus,M.G., Hsieh,L.C., Janusz,M.J., Peng,S.X., Branch,T.M., King,S.L., Taiwo,Y.O., & Mieling,G.E. Potent and selective carboxylic acid-based inhibitors of matrix metalloproteinases. *J. Med. Chem.* **44**, 2499-2502 (2001).
83. Auge,F., Hornebeck,W., Decarme,M., & Laronze,J.Y. Improved gelatinase A selectivity by novel zinc binding groups containing galardin derivatives. *Bioorg. Med. Chem. Lett.* **13**, 1783-1786 (2003).
84. Campbell,D.A., Xiao,X.Y., Harris,D., Ida,S., Mortezaei,R., Ngu,K., Shi,L., Tien,D., Wang,Y., Navre,M., Patel,D.V., Sharr,M.A., DiJoseph,J.F., Killar,L.M., Leone,C.L., Levin,J.I., & Skotnicki,J.S. Malonyl alpha-mercaptoketones and alpha-mercaptoalcohols, a new class of matrix metalloproteinase inhibitors. *Bioorg. Med. Chem. Lett.* **8**, 1157-1162 (1998).
85. Brown,M., Bernardo,M.M., Li,Z., Kotra,L.P., Tanaka,Y., & Mobashery,S. Potent and Selective Mechanism-Based Inhibition of Gelatinases. *J. Am. Chem. Soc.* **122**, 6799-6800 (2000).
86. Bernardo,M.M., Brown,S., Li,Z.H., Fridman,R., & Mobashery,S. Design, synthesis, and characterization of potent, slow-binding inhibitors that are selective for gelatinases. *J. Biol. Chem.* **277**, 11201-11207 (2002).

87. Dunten,P., Kammlott,U., Crowther,R., Levin,W., Foley,L.H., Wang,P., & Palermo,R. X-ray structure of a novel matrix metalloproteinase inhibitor complexed to stromelysin. *Protein Sci.* **10**, 923-926 (2001).
88. Brandstetter,H., Grams,F., Glitz,D., Lang,A., Huber,R., Bode,W., Krell,H.W., & Engh,R.A. The 1.8-Å crystal structure of a matrix metalloproteinase 8-barbiturate inhibitor complex reveals a previously unobserved mechanism for collagenase substrate recognition. *J. Biol. Chem.* **276**, 17405-17412 (2001).
89. Gall,A.L., Ruff,M., Kannan,R., Cuniasse,P., Yiotakis,A., Dive,V., Rio,M.C., Basset,P., & Moras,D. Crystal structure of the Stromelysin-3 (MMP-11) catalytic domain complexed with a phosphinic inhibitor mimicking the transition-state. *J. Mol.* **307**, 577-586 (2001).
90. Gavuzzo,E., Pochetti,G., Mazza,F., Gallina,C., Gorini,B., D'Alessio,S., Pieper,M., Tschesche,H., & Tucker,P.A. Two crystal structures of human neutrophil collagenase, one complexed with a primed- and the other with an unprimed-side inhibitor: implications for drug design. *J. Med. Chem.* **43**, 3377-3385 (2000).
91. Devel,L., Rogakos,V., David,A., Makaritis,A., Beau,F., Cuniasse,P., Yiotakis,A., & Dive,V. Development of selective inhibitors and substrate of matrix metalloproteinase-12. *J. Biol. Chem.* **281**, 11152-11160 (2006).
92. Puerta,D.T., Lewis,J.A., & Cohen,S.M. New beginnings for matrix metalloproteinase inhibitors: Identification of high-affinity zinc-binding groups. *J. Am. Chem. Soc.* **126**, 8388-8389 (2004).
93. Mannino,C., Nievo,M., Machetti,F., Papakyriakou,A., Fragai,M., & Guarna,A. Synthesis of bicyclic molecular scaffolds (BTAA) for the development of new matrix metalloproteinases inhibitors. *Bioorg. Med. Chem.* **14**, 7392-7403 (2006).
94. Engel,C.K., Pirard,B., Schimanski,S., Kirsch,R., Habermann,J., Klingler,O., Schlotte,V., Weithmann,K.U., & Wendt,K.U. Structural basis for the highly selective inhibition of MMP-13. *Chem. Biol.* **12**, 181-189 (2005).
95. Johnson,A.R., Pavlovsky,A.G., Ortwine,D.F., Prior,F., Man,C.F., Bornemeier,D.A., Banotai,C.A., Mueller,W.T., McConnell,P., Yan,C., Baragi,V., Lesch,C., Roark,W.H., Wilson,M., Datta,K., Guzman,R., Han,H.K., & Dyer,R.D. Discovery and characterization of a novel inhibitor of matrix metalloproteinase-13 that reduces cartilage damage in vivo without joint fibroplasia side effects. *J. Biol. Chem.* **282**, 27781-27791 (2007).
96. Li,J.J., Naha,J., Johnson,A.R., Bunker,A., O'Brien,P., Yue,W.S., Ortwine,D.F., Man,C.F., Baragi,V., Kilgore,K., Dyer,R.D., & Han,H.K. Quinazolinones and pyrido[3,4-d]pyrimidin-4-ones as orally active and specific matrix metalloproteinase-13 inhibitors for the treatment of osteoarthritis. *J. Med. Chem.* **51**, 835-841 (2008).
97. Pirard,B. Insight into the structural determinants for selective inhibition of matrix metalloproteinases. *Drug Discov. Today* **12**, 640-646 (2007).
98. von Itzstein,M., Wu,W.Y., Kok,G.B., Pegg,M.S., Dyason,J.C., Jin,B., Van Phan,T., Smythe,M.L., White,H.F., & Oliver,S.W. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **363**, 418-423 (1993).
99. Habeck,M. FDA licences imatinib mesylate for CML. *Lancet Oncol.* **3**, 6 (2002).
100. Moy,F.J., Chanda,P.K., Chen,J., Cosmi,S., Edris,W., Levin,J.I., Rush,T.S., Wilhelm,J., & Powers,R. Impact of mobility on structure-based drug design for the MMPs. *J. Am. Chem. Soc.* **124**, 12658-12659 (2002).
101. Lovejoy,B., Welch,A.R., Carr,S., Luong,C., Broka,C., Hendricks,R.T., Campbell,J.A., Walker,K.A.M., Martin,R., Van Wart,H., & Browner,M.F. Crystal structures of MMP-1 and -13 reveal the structural basis for selectivity of collagenase inhibitors. *Nat. Struct. Biol.* **6**, 217-221 (1999).

102. Bhaskaran,R., Palmier,M.O., Bagegni,N.A., Liang,X.Y., & Van Doren,S.R. Solution structure of inhibitor-free human metalloelastase (MMP-12) indicates an internal conformational adjustment. *J. Mol. Biol.* **374**, 1333-1344 (2007).
103. Ishima,R. & Torchia,D.A. Protein dynamics from NMR. *Nature Struct. Biol.* **7**, 740-743 (2000).
104. Kay,L.E. Protein dynamics from NMR. *Nature Struct. Biol.* **5**, 513-517 (1998).
105. Lindorff-Larsen,K., Best,R.B., DePristo,M.A., Dobson,C.M., & Vendruscolo,M. Simultaneous determination of protein structure and dynamics. *Nature* **433**, 128-132 (2005).
106. Fischer,M.W.F., Zeng,L., Majumdar,A., & Zuiderweg,E.R.P. Characterizing semilocal motions in proteins by NMR relaxation studies. *Proc. Natl. Acad. Sci. USA* **95**, 8016-8019 (1998).
107. Mittermaier,A. & Kay,L.E. New tools provide new insights in NMR studies of protein dynamics. *Science* **312**, 224-228 (2006).
108. Fragai,M., Luchinat,C., & Parigi,G. "Four-dimensional" protein structures: examples from metalloproteins. *Acc. Chem. Res.* **39**, 909-917 (2006).
109. Bertini,I., Calderone,V., Cosenza,M., Fragai,M., Lee,Y.-M., Luchinat,C., Mangani,S., Terni,B., & Turano,P. Conformational variability of MMPs: beyond a single 3D structure. *Proc. Natl. Acad. Sci. USA* **102**, 5334-5339 (2005).
110. Miller,E.J., Harris,E.D., Jr., Chung,E., Finch,J.E., Jr., McCroskery,P.A., & Butler,W.T. Cleavage of Type II and III collagens with mammalian collagenase: site of cleavage and primary structure at the NH<sub>2</sub>-terminal portion of the smaller fragment released from both collagens. *Biochemistry* **15**, 787-792 (1976).
111. Birkedal-Hansen,H., Taylor,R.E., Bhowan,A.S., Katz,J., Lin,H.Y., & Wells,B.R. Cleavage of bovine skin type III collagen by proteolytic enzymes. Relative resistance of the fibrillar form. *J. Biol. Chem.* **260**, 16411-16417 (1985).
112. Ottil,J., Battistuta,R., Pieper,M., Tschesche,H., Bode,W., Kuhn,K., & Moroder,L. Design and synthesis of heterotrimeric collagen peptides with a built-in cystine-knot. Models for collagen catabolism by matrix-metalloproteases. *FEBS Lett.* **398**, 31-36 (1996).
113. Kramer,R.Z., Bella,J., Mayville,P., Brodsky,B., & Berman,H.M. Sequence dependent conformational variations of collagen triple-helical structure. *Nat. Struct. Biol.* **6**, 454-457 (1999).
114. Orgel,J.P., Irving,T.C., Miller,A., & Wess,T.J. Microfibrillar structure of type I collagen in situ. *Proc. Natl. Acad. Sci. U. S. A* **103**, 9001-9005 (2006).
115. Ottil,J., Gabriel,D., Murphy,G., Knauper,V., Tominaga,Y., Nagase,H., Kroger,M., Tschesche,H., Bode,W., & Moroder,L. Recognition and catabolism of synthetic heterotrimeric collagen peptides by matrix metalloproteinases. *Chemistry & Biology* **7**, 119-132 (2000).
116. Yu,Y.C., Roontga,V., Daragan,V.A., Mayo,K.H., Tirrell,M., & Fields,G.B. Structure and dynamics of peptide-amphiphiles incorporating triple-helical proteinlike molecular architecture. *Biochemistry* **38**, 1659-1668 (1999).
117. Chung,L.D., Dinakarandian,D., Yoshida,N., Lauer-Fields,J.L., Fields,G.B., Visse,R., & Nagase,H. Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis. *EMBO J.* **23**, 3020-3030 (2004).
118. Li,J., Brick,P., O'Hare,M.C., Skarzynski,T., Lloyd,L.F., Curry,V.A., Clark,I.M., Bigg,H.F., Hazleman,B.L., & Cawston,T.E. Structure of full-length porcine synovial collagenase reveals a C-terminal domain containing a calcium-linked, four-bladed beta-propeller. *Structure* **3**, 541-549 (1995).
119. Guex,N. & Peitsch,M.C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723 (1997).

120. Williams,K.E. & Olsen,D.R. Matrix metalloproteinase-1 cleavage site recognition and binding in full-length human type III collagen. *Matrix Biol.* **28**, 373-379 (2009).
121. Jozic,D., Bourenkov,G., Lim,N.H., Visse,R., Nagase,H., Bode,W., & Maskos,K. X-ray structure of human proMMP-1 - New insights into procollagenase activation and collagen binding. *J. Biol. Chem.* **280**, 9578-9585 (2005).
122. Piccard,H., Van den Steen,P.E., & Opdenakker,G. Hemopexin domains as multifunctional liganding modules in matrix metalloproteinases and other proteins. *J. Leukoc. Biol.* **81**, 870-892 (2007).
123. Monaco,S., Gioia,M., Rodriguez,J., Fasciglione,G.F., Di Pierro,D., Lupidi,G., Krippahl,L., Marini,S., & Coletta,M. Modulation of the proteolytic activity of matrix metalloproteinase-2 (gelatinase A) on fibrinogen. *Biochem. J.* **402**, 503-513 (2007).
124. Roeb,E., Schleinkofer,K., Kernebeck,T., Pötsch,S., Jansen,B., Behrmann,I., Matern,S., & Grötzinger,J. The MMP-9 hemopexin domain is a novel gelatin binding domain and acts as an antagonist. *J. Biol. Chem.* **277**, 50326-50332 (2002).
125. Morgunova,E., Tuuttila,A., Bergmann,U., Isupov,M., Lindqvist,Y., Schneider,G., & Tryggvason,K. Structure of human pro-matrix metalloproteinase-2: activation mechanism revealed. *Science* **284**, 1667-1670 (1999).
126. Steffensen,B., Wallon,U.M., & Overall,C.M. Extracellular matrix binding properties of recombinant fibronectin type II-like modules of human 72-kDa gelatinase/type IV collagenase. High affinity binding to native type I collagen but not native type IV collagen. *J. Biol. Chem.* **270**, 11555-11566 (1995).
127. Xu,X., Wang,Y., Lauer-Fields,J.L., Fields,G.B., & Steffensen,B. Contributions of the MMP-2 collagen binding domain to gelatin cleavage. Substrate binding via the collagen binding domain is required for hydrolysis of gelatin but not short peptides. *Matrix Biol.* **23**, 171-181 (2004).
128. Shipley,J.M., Doyle,G.A., Fliszar,C.J., Ye,Q.Z., Johnson,L.L., Shapiro,S.D., Welgus,H.G., & Senior,R.M. The structural basis for the elastolytic activity of the 92-kDa and 72-kDa gelatinases. Role of the fibronectin type II-like repeats. *J. Biol. Chem.* **271**, 4335-4341 (1996).
129. Gronski,T.J., Martin,R.L., Kobayashi,D.K., Walsh,B.C., Holman,M.C., Huber,M., VanWart,H., & Shapiro,S.D. Hydrolysis of a broad spectrum of extracellular matrix proteins by human macrophage elastase. *J. Biol. Chem.* **272**, 12189-12194 (1997).
130. Mecham,R.P., Broekelmann,T.J., Fliszar,C.J., Shapiro,S.D., Welgus,H.G., & Senior,R.M. Elastin degradation by matrix metalloproteinases. Cleavage site specificity and mechanisms of elastolysis. *J. Biol. Chem.* **272**, 18071-18076 (1997).
131. Liu,M., Sun,H., Wang,X., Koike,T., Mishima,H., Ikeda,K., Watanabe,T., Ochiai,N., & Fan,J. Association of increased expression of macrophage elastase (matrix metalloproteinase 12) with rheumatoid arthritis. *Arthritis Rheum.* **50**, 3112-3117 (2004).
132. Taddese,S., Weiss,A.S., Neubert,R.H., & Schmelzer,C.E. Mapping of macrophage elastase cleavage sites in insoluble human skin elastin. *Matrix Biol.* **27**, 420-428 (2008).
133. Taddese,S., Weiss,A.S., Jahreis,G., Neubert,R.H., & Schmelzer,C.E. In vitro degradation of human tropoelastin by MMP-12 and the generation of matrikines from domain 24. *Matrix Biol.* **28**, 84-91 (2009).
134. Schwartz,E., Feinberg,E., Lebwohl,M., Mariani,T.J., & Boyd,C.D. Ultraviolet radiation increases tropoelastin accumulation by a post-transcriptional mechanism in dermal fibroblasts. *J. Invest Dermatol.* **105**, 65-69 (1995).

## **2. Materials and Methods**

## 2.1 NMR

### 2.1.1 Protein assignment experiments

The experiments for the protein assignment, mobility measurements and titration experiments of the isolated CAT and HPX domain and of the full length proteins MMP1 and MMP12 were performed on protein samples at concentrations ranging between 0.2 and 0.7 mM. Sample conditions were usually 0.1 M of NaCl for ionic strength, 10 mM of CaCl<sub>2</sub> and 0.1 mM of ZnCl<sub>2</sub> using a tris-HCl buffer at pH 7.2 .

All NMR experiments were performed at 298K, 306K and 310K and acquired on Bruker AVANCE900, AVANCE800, AVANCE700 AVANCE 600 and DRX 500 spectrometers. All instruments are equipped with triple resonance CRYO-probes.

Experiments for backbone resonance assignment of HPX-MMP12 and HPX-MMP1 : HSQC<sup>1,2,3,4</sup>, HNCA<sup>5</sup>, CBCA(CO)NH<sup>6</sup> and HNCACB<sup>7,8</sup> were performed at 900MHz for HPX-MMP12 and 500 MHz for HPX-MMP1. For HPX-MMP12, spectra were recorded at 298K while for HPX-MMP1, temperature was set to 310K to prevent aggregation.

The assignment of the 2H-, 15N-, and 13C-enriched FL-MMP1 was obtained by the analysis of 1H-15N TROSY<sup>9</sup> TROSY-HNCA<sup>10</sup> and TROSY-HNCACB<sup>11,12</sup> performed on an 800 MHz spectrometer at 310K. A 13C decoupling sequence was used on 1H-15N 2D TROSY experiment and a 2H decoupling sequence was used for the 3-dimensional triple resonance experiments. The protein assignment and the mobility measurements for both FL-MMP12 and FL-MMP1 were performed on the NNGH-inhibited and E219A mutant proteins, necessary conditions to provide high stability by preventing the self-hydrolysis activity.

The assignment of the aliphatic side-chain resonances of HPX-MMP12 was performed through the analysis of 3D (H)CCH-TOCSY<sup>13</sup> spectra at 500 MHz, together with 3D 15N- and 13C-NOESY-HSQC<sup>14,15,16,17,18</sup> spectra at 900 MHz. These experiments were performed at 298K.

J<sub>HNH</sub> coupling constants were determined through the HNHA<sup>19,20</sup> experiment at 500 MHz. Backbone dihedral  $\phi$  angles were independently derived from J<sub>HNH</sub> coupling constants through the appropriate Karplus equation<sup>21</sup>. Backbone dihedral  $\psi$  angles for residue i-1 were also determined.

The protein assignment and the mobility measurements on FL-MMP12 were performed on the NNGH-inhibited, cadmium(II)- substituted Phe171Asp/Glu219Ala mutant<sup>22</sup>, due to its high stability to the self-hydrolysis.

### 2.1.2 Residual dipolar coupling

Residual dipolar couplings (RDCs) have been measured on FL-MMP12 in the presence of an external orienting medium constituted by a binary mixture of C<sub>12</sub>E<sub>5</sub> (penta-ethyleneglycoldodecylether, Fluka) and neat n-hexanol (Fluka), with a molar ratio C<sub>12</sub>E<sub>5</sub>/n-hexanol of 0.96 and with a C<sub>12</sub>E<sub>5</sub>/water ratio of 5wt%.<sup>23</sup> One-bond 1H 15N coupling constants were measured at 298K and 900MHz by using the IPAP method.<sup>24</sup> Two-hundred fifty-nine RDC values could be measured that ranged from 46 to 25Hz. Of them, only those RDC values corresponding to residues experiencing neither mobility nor large RMSD (140 residues, mostly in  $\alpha$  or  $\beta$  secondary structures) have been used for structure calculations<sup>25</sup> and to investigate the reciprocal mobility of the two domains.

### 2.1.3 Pseudo-contact shift

In order to visualise pseudo-contact shifts (PCS)<sup>26</sup> from a FL-MMP12 protein sample in presence of a paramagnetic cobalt ion in the active site, 2D HSQC and TROSY experiments were performed. The sample was prepared by substituting the natural catalytic zinc metal ion of the enzyme by a cobalt ion in a dialyse membrane. A reference spectrum from a sample containing the zinc ion was also acquired at the 900 MHz. Comparison between reference spectrum and spectrum obtained in paramagnetic condition, allowed to visualise PCS.

Similar experiments were performed with a CAT-MMP12 sample containing a cobalt ion and another CAT-MMP12 with a zinc ion. These experiments using the isolated CAT domains were performed to compare the PCS observed from the CAT domain alone and the FL protein.

### 2.1.4 Relaxation experiments

#### T1, T2 and NOE

Mobility measurements on the catalytic domain were performed on the NNGH-inhibited, zinc (II) form of the Phe171Asp mutant.<sup>22</sup>

Experiments for the determination of <sup>15</sup>N longitudinal and transverse relaxation rates,<sup>15</sup> and <sup>1</sup>H-<sup>15</sup>N NOE<sup>27</sup> were performed at 298 K and 700 MHz on <sup>15</sup>N-enriched samples of FL-MMP12 and isolated domains. The <sup>15</sup>N longitudinal relaxation rates (*R*<sub>1</sub>) were measured using a sequence modified to remove cross correlation effects during the relaxation delay.<sup>28</sup> Inversion recovery times ranging between 2.5 and 3000 ms, with a recycle delay of 3.5 s, were used for the experiments. The <sup>15</sup>N transverse relaxation rates (*R*<sub>2</sub>) were measured using a Carr-Purcell-Meiboom-Gill (CPMG) sequence<sup>29</sup> with delays ranging between 8.5 and 237.4 ms for the CAT domain, between 8.5 and 203.5 ms for the HPX domain, and finally between 8.5 and 135.7 ms for the FL-MMP12 protein with a refocusing delay of 450 μs.

Similar relaxation experiments were performed with CAT, HPX and FL-MMP1 samples at the 700 MHz spectrometer with a temperature of 306K.

#### Paramagnetic Relaxation Enhancement Measurements

The paramagnetic relaxation enhancements of the backbone NH protons<sup>30</sup> were obtained by adding a stock water solution (50mM) of Gd(DTPA\_BMA) to a solution of FL-MMP1 (0.2 mM protein in a buffer containing 20mM Tris, pH 7.2, 10mM CaCl<sub>2</sub>, 0.3 M NaCl, 0.2 M acetohydroxamic acid, 1mM ZnCl<sub>2</sub>, and 3mM NNGH) up a final concentration of 1.4 mM. Experiments were also performed at 700 MHz at 306°K.

NH proton longitudinal relaxation times were measured through an inversion recovery HSQC sequence obtained by introducing a <sup>1</sup>H 180° pulse followed by a variable delay  $\tau$  in front of a standard <sup>1</sup>H-<sup>15</sup>N-HSQC sequence. T1 values were obtained from a series of spectra obtained with the following  $\tau$  values (in ms): 20, 50, 80, 100, 200, 400, 500, 800, 1000, 1400, 2000, 2200, 3000, and 3500. A recycle delay of 4 seconds was inserted between two scans. The measurement was repeated on the free protein and after each addition of the paramagnetic complex. Peak volumes as a function of the  $\tau$  value were fitted to a monoexponential recovery with a three-parameter fit.



### 2.1.5 NMR titration experiments

#### MMP12 Interaction studies with elastin

Full-length MMP12, its CAT domain and the HPX domain were expressed and purified as described previously.<sup>31</sup> The replacement of the catalytic zinc with the inactive cadmium ion was performed by exhaustive dialysis against a buffer containing 20 mM Tris pH 7.2, 10 mM CaCl<sub>2</sub>, 0.15 M NaCl, 0.2 M acetohydroxamic acid (AHA) and 0.3 mM of CdCl<sub>2</sub>. The acetohydroxamic acid was removed during the last dialysis. Soluble elastin from bovine neck ligament was purchased from Sigma–Aldrich. Lyophilised elastin was dissolved in the final buffer (20 mM Tris pH 7.2, 10 mM CaCl<sub>2</sub>, 0.15M NaCl and 0.3 mM of CdCl<sub>2</sub>) and extensively dialyzed before the addition to the protein solution.

In order to study the interaction in solution by NMR between MMP12 and elastin, HSQC, TROSY and FAST HMQC<sup>32</sup> 2D experiments were performed at the 700 MHz and 900 MHz. The reference spectra were acquired on the protein alone and other titration spectra were recorded after each addition of elastin. Experiments were performed for the FL-MMP12, CAT domain and HPX domain alone and in presence of elastin fragments.

Soluble elastin (0.7 mg) was added to 0.5 mL of FL-MMP12 (0.35 mM). The same amount of soluble elastin was added to 0.5 mL of isolated CAT and HPX domains at a concentration of 0.3 mM. The final concentration of soluble elastin in solution, calculated as tropoelastin (MW~65 kDa), was estimated to be  $2.2 \times 10^{-2}$  mM. However, taking into account the presence of several cleavage sites in the substrate (up to 86 in tropoelastin and up to 36 in insoluble elastin<sup>33,34</sup>), the concentration of binding sites for the catalytic domain in the samples may range from 0.8 up to 1.9 mM, that is, in large stoichiometric excess. The heterogeneous composition of soluble elastin and the presence of several binding sites, plausibly with different affinities, prevent the calculation of any dissociation constant.

#### MMP1 interaction studies with a triple helical peptide (THP) model of collagen

For MMP1 protein, experiments were performed using FL-MMP1 in absence and in presence of the triple helical peptide (THP) model of collagen<sup>35</sup>. The samples were prepared and concentrated till 0.2 mM in a buffer containing 0.1 M NaCl ; 10 mM CaCl<sub>2</sub> ; 1 mM CdCl<sub>2</sub> and 20 mM tris-HCl at pH 7.2 . TROSY spectra were acquired at the 900 MHz spectrometer at a temperature of 310K. Peptide was added in solution to reach a final concentration of 0.08 mM. Due to a severe decrease in peak intensity, it was impossible to increase the concentration of the peptide.

In order to assign the unlabelled THP, two dimensional H-H TOCSY<sup>36</sup> spectra were also acquired at 700 MHz spectrometer at a temperature of 310K. First, 1 mg of lyophilised THP (MW ≈ 10 kDa) was dissolved into water to obtain a reference spectrum for NMR assignment at a concentration of 0.5 mM. Then another mg of THP was dissolved into a buffer tris-HCl pH 7.2 to report its assignment in the buffer condition of interaction.

To visualise chemical shift changes on the HPX domain, NMR titration was performed acquiring HSQC and TROSY spectra at the 900 MHz spectrometer. A reference spectrum was acquired with a protein sample at 0.2 mM. Several amounts of peptide model THP were added to the HPX-MMP1 sample to reach final peptide concentrations respectively of 0.04, 0.08, 0.15, 0.23, 0.30 and 0.40 mM.

TOCSY spectra were recorded at 700 MHz spectrometer from HPX domain 0.2 mM (control), THP peptide 0.5 mM and HPX-THP complex from the previous titration (0.2 mM : 0.4 mM). Experiments were performed at 310°K and 298°K.

## 2.2 Assignment strategy

All spectra were processed with the Bruker TOPSPIN software Packages and analyzed by the program CARA (Computer Aided Resonance Assignment, ETH Zürich).<sup>37</sup>

Assignment can be divided into three main steps. The first step consisting in identifying the amide proton and nitrogen located in the backbone of the protein is called backbone assignments.

The backbone resonance assignment of HPX-MMP12 and HPX-MMP1 was obtained by the analysis of HNCA, HNCOCACB, and CBCA(CO)NH. The assignment of the aliphatic side-chain resonances was performed through the analysis of 3D (H)CCH-TOCSY, together with 3D <sup>15</sup>N- and <sup>13</sup>C-NOESY-HSQC.

The backbone assignment of the <sup>2</sup>H-, <sup>15</sup>N-, and <sup>13</sup>C-enriched FL-MMP1 was obtained by the analysis TROSY-HNCA, TROSY-HNCACB, TROSY-HNCOCACB.

The technique requires comparing the 2D and 3D spectra in order to associate the different carbon resonances which belong to each amide proton.

Once these are identified, it becomes possible to find connexions between the amide protons through alpha and beta carbons. The fragments obtained are then analysed in order to be assigned inside the sequence of the protein. Each protein has a unique assignment corresponding to its 3D structure. However, proteins which belong to the same family may have several resonances close to each other. In the example of HPX-MMP1 assignment, its sequence was aligned with the one of HPX-MMP12. Then, the available PDB structures have been superimposed and eventually, NMR assignment of HPX-MMP12 was used as reference assignment to help finding HPX-MMP1 assignment. In addition to the comparison with HPX-MMP12, SPARTA<sup>38</sup> and PROSHIFT<sup>39</sup> programs were used to obtain a chemical shift prediction from X-ray or NMR structure. The programme SPARTA is based on the relation between dihedral angles found in structures contained in the PDB database for the backbone and the backbone assignment chemical shifts deposited on BMRB database. Then, the programme calculates the best agreement of probable chemical shifts with the submitted structure. PROSHIFT programme has also been used. It also starts from the PDB structure but uses a different algorithm based on artificial neurones which combines entry information and provides the best compromise.

Backbone assignment is mainly used to study the structure and mobility of the protein. In addition, backbone assignment is used to work on RDCs and PCS. It is also used to study the interaction between the protein and a partner.

After backbone assignment the two following steps of assignment will depend strongly on it. Second step is the side-chain assignment used to identify carbons and protons of each residue which belong to the same spin system starting from already assigned C $\alpha$  and C $\beta$ . Some side-chain H $\alpha$  and H $\beta$  protons are also required to start the side-chain assignment. These resonances are also provided by the use on the <sup>15</sup>N 3D HNH NOESY spectra. The third step consists in finding connections between protons by studying signals observed on NOESY spectra. These NOE cross peaks can be integrated and the volume obtained is correlated to the distances between atoms. This information is required to calculate the structures (see chapter related on structure calculation).

## 2.3 Mobility studies from relaxation data

To calculate longitudinal and transverse relaxation, a reference peaklist was obtained by importing the protein amide proton assignment on the reference spectra of mobility (in generally the shortest relaxation time). Then, peaks are centred in the maximum of intensity. A normalisation is realised to obtain a homogeneous integration of the different peaks. All the peaks are integrated in all the different spectra. Then a plot is realised by reporting intensity (volume) as a function of time. Fitting experimental data is obtained from the equation :

$$\mathbf{f(t) = I_0 \times e^{(-Rt)} + A_0}$$

Where  $f(t)$  is the volume calculated,  $t$  is the variable time in seconds used and  $R$  is the relaxation constant  $R1$  or  $R2$  in ( $s^{-1}$ ) which has to be determined.  $I_0$  and  $A_0$  are constants for the fitting of the curve. The  $R1$  and  $R2$  parameters provide indications for the regional mobility for the molecule.  $R1$  and  $R2$  parameter can be fitted by using the exponential decay function in the software origin.

In addition to  $R1$ ,  $R2$  determination, NOE experiment can provide NOE data by a splitting on the spectra (NOE and non-NOE). Non-NOE/NOE ratio calculated for each residue provides information about the local mobility.

Estimates of  $R1$  and  $R2$  values for CAT and HPX and FL-MMP12 and for CAT, HPX and FL-MMP1 were determined by using the program HydroNMR<sup>40</sup> which simulates conditions of magnetic field and temperature from the three-dimensional structures of the proteins taken as input for the programme with respectively the X-ray structure of the FL-MMP1<sup>41</sup> and the corresponding model of FL-MMP12 protein.

Regarding paramagnetic relaxation enhancement measurements in the presence of the paramagnetic probe, peak intensity volumes are integrated in same the way as for the determination  $^{15}N$  longitudinal and transverse  $R1$  and  $R2$  relaxation rates by preparing a peaklist.

Then  $R$  parameter is obtained from the equation

$$\mathbf{f(t) = I_0 \times (1 - 2 \times e^{(-Rt)}) + A_0}$$

Where  $f(t)$  is the volume calculated,  $t$  is the variable time in seconds used and  $R$  is the relaxation constant in ( $s^{-1}$ ) which has to be determined.  $I_0$  and  $A_0$  are constants for the fitting of the curve.

## 2.4 Structure calculations

The first step consists in generating an upper limit distance by calculating NOE cross peaks volumes. Caliba program was used for this purpose. A calibration is required because the NOE signals observed on the spectra depend on the nature of the cross-peaks. For example, intensity measured may be influenced in case of degenerated protons or if these protons are directly connected through bonds.

From the intensity ratio of the  $dN(i-1,i)$  and  $dN(i,i)$  NOEs present on the  $^{15}N(i)$  plane of residue  $i$  obtained from the  $^{15}N$ -edited NOESY-HSQC spectrum 3D  $^{15}N$ -and  $^{13}C$ -enriched NOESY-HSQC cross-peak intensities were integrated using the integration routine implemented in CARA and converted into interatomic upper distance limits by CALIBA<sup>42</sup> programme.

A solution structure can be calculated using also torsion angles located on the backbone. As mentioned in the paragraph related to NMR experiment, angles can also be determined by using the properties of Karplus Curve from HNHA experiments. Else, Chemical Shift Index (CSI)<sup>43</sup> method can predict the secondary structure of residues from the chemical shifts of the backbone atoms. TALOS<sup>44</sup> programme can be used to predict a torsion angle range from chemicals shifts measured on to three consecutives residues. In addition, TALOS can approximate angles for residues located in loops or turns of a protein.

Nevertheless, The  $J_{HNH}$  coupling constant method based on experiment data remains more reliable rather than CSI or TALOS predictions based on statistical data. However, a comparison between experimentally defined secondary structure and prediction can provide a confirmation of the data.

DYANA<sup>45</sup> programme was used to calculate the NMR solution structure of the protein by combining distances, angles (restraints) and RDCs. Each calculation generates a family of structures on a pdb format and an output file with an overview containing the target function of each conformer with a list of violations and the RMSD (root mean square deviation) of the ensemble. The pdb file can be displayed by using molmol<sup>46</sup> software initially developed to analyse NMR ensemble structure from calculation. The target function depends on the agreement between constraints and calculated structure. By checking the list of violations, mistakes can be indentified to be corrected in order to improve the structure quality. RMSD indicate the structure precision considering convergence of the data. Then, RDCs constraints complementary to distance and angle constraints generate a protein tensor.

DYANA was run to calculate a family of 1600 structures of the isolated HPX domain starting from randomly generated conformers in 20000 annealing steps. The family was energy-minimized by iterative cycles of DYANA with the program FANTAORIENT.<sup>47</sup> Structure calculation statistics and the structural quality were evaluated using the program PROCHECK\_NMR<sup>48</sup> which in particular displays the Ramachandran plot<sup>49</sup>. Most of the residues must have their respective torsion angle in the core and allowed regions rather than in generously and disallowed regions.

The structure obtained by DYANA calculation can be also minimised to decrease total energy by using AMBER<sup>50</sup> programme and then re-evaluated by ProCHECK-NMR

In addition to RDCs, complementary constraints can be obtained by the use of PCS. PCS observed on the spectra are conversely proportional to the distance between the paramagnetic ion and a given amino-acid located in the range of paramagnetic effects. Interestingly, a paramagnetic metal ion can provide long range constraints with distances higher than 10 Å.

## 2.5 Docking Calculation

Haddock<sup>51</sup> programme online was used to calculate the interaction model between MMP12 and elastin fragments and between MMP1 and the collagen peptide model. The first step consisted in selecting active residues in order to drive the docking on the base on their chemical shift perturbation from NMR experiments and solvent accessibility. The two latter conditions should be respected. Regarding chemical shift perturbation, a threshold needs to be defined. For example, residues with chemical shift changes higher than the average value plus the standard deviation value can be selected. Second, relative accessibility for a defined residue should exceed 50%. Naccess<sup>52</sup> programme calculates protein surface from a pdb file and provides information about accessibility of each individual residue. Once active residues have been defined for the two partners of the interaction, the next step consists in selecting the flexible fragments. These flexible fragments for docking were chosen around the active residue to allow local conformation change to improve the energy binding.

At the end of the calculation run, results are classified according to an energy score and grouped into clusters on the basis of an RMSD cut-off. Output files for energy and RMSD are provided by the programme.

### Interaction model of a MMP12 – Elastin fragments complex

Regarding interaction between MMP12 and elastin fragments, selection of active residues was done from chemical shift changes during titration and intensity decrease of some peaks from the CAT domain (see results section). From elastin, only the information about the cleavage site sequences is available<sup>33</sup> which allow binding the CAT domain at its active site close to the metal ion to elastin peptide fragments which can be randomly generated by DYANA programme. Information about possible binding of the HPX to elastin has not been reported to date.

### Interaction model of a MMP1 - triple helical peptide (THP) model of collagen complex

For MMP1, the selection was done for both partners of the docking (MMP1 and THP). Preliminary docking was performed from results obtained on separated domains. Atom coordinates of MMP1 were taken from the FL X-Ray structure (pdb entry 2CLT).<sup>41</sup>

For the THP, a model was generated according to the crystal collagen peptide structures already known (see introduction p. 14).

Active residues on HPX domain were selected from the titration results in the presence of THP by HSQC and TROSY experiments. Reciprocal active residues on THP were taken from chemical shift changes on the TOCSY spectra.

Active residues on CAT domain were taken from chemical shift changes from titration of the FL-protein by THP. Additional distances were used from an X-ray structure of CAT-MMP1 in complex with a single chain collagen fragment. Reciprocal active residues from the THP were selected around the cleavage site.

In order to obtain the interaction model of the FL-MMP1 with THP, two strategies were used : one consisted in fitting the two domains docking models on THP. Then, the linker peptide was rebuilt using modeller<sup>53</sup> and Haddock was rerun on the rebuilt FL-MMP1. A complementary approach consisted in starting from the X-ray structure. A first docking step consisted in binding the HPX domain to the THP. Then, CAT domain was moved around THP using a flexible linker parameter by iterative docking steps to finally reach the cleavage site.

## 2.6. Reference List

1. Bodenhausen, G. & Ruben, D.J. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69**, 185-188 (1980).
2. Piotto, M., Saudek, V., & Sklenar, V. Gradient-tailored excitation for single quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **2**, 661-666 (1992).
3. Sklenar, V., Piotto, M., Leppik, R., & Saudek, V. Gradient-tailored water suppression for  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments optimized to retain full sensitivity. *J. Magn. Reson. Ser. A* **102**, 241-245 (1993).
4. Mori, S., Abeygunawardana, C., Johnson, M.O., & van Zijl, P.C. Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J. Magn Reson. B* **108**, 94-98 (1995).
5. Grzesiek, S. & Bax, A. Improved 3D Triple-Resonance NMR Techniques Applied to a 31 KDa Protein. *J. Magn. Reson.* **96**, 432-440 (1992).
6. Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., & Kay, L.E. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by  $^{15}\text{N}$  NMR relaxation. *Biochemistry* **33**, 5984-6003 (1994).
7. Grzesiek, S. & Bax, A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.* **99**, 201-207 (1992).
8. Muhandiram, D.R. & Kay, L.E. Gradient-enhanced triple resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. Ser. B* **103**, 203-216 (1994).
9. Czisch, M. & Boelens, R. Sensitivity enhancement in the TROSY experiment. *J. Magn Reson.* **134**, 158-160 (1998).
10. Salzmann, M., Pervushin, K., Wider, G., Senn, H., & Wüthrich, K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. Natl. Acad. Sci. USA* **95**, 13585-13590 (1998).
11. Salzmann, M., Wider, G., Pervushin, K., Senn, H., & Wüthrich, K. TROSY-type Triple-Resonance Experiments for Sequential NMR Assignments of Large Proteins. *J. Am. Chem. Soc.* **121**, 844-848 (1999).
12. Eletsky, A., Kienhofer, A., & Pervushin, K. TROSY NMR with partially deuterated proteins. *J. Biomol. NMR* **20**, 177-180 (2001).
13. Kay, L.E., Xu, G.Y., Singer, A.U., Muhandiram, D.R., & Forman-Kay, J.D. A gradient-enhanced HCCH-TOCSY experiment for recording side-chains  $^1\text{H}$  and  $^{13}\text{C}$  correlations in  $\text{H}_2\text{O}$  samples of proteins. *J. Magn. Reson. Ser. B* **101**, 333-337 (1993).
14. Zhang, O., Kay, L.E., Olivier, J.P., & Forman-Kay, J.D. Backbone  $^1\text{H}$  and  $^{15}\text{N}$  resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR* **4**, 845-858 (1994).
15. Palmer, A.G., III, Rance, M., & Wright, P.E. Intramolecular motions of a zinc finger DNA-binding domain Xfin characterized by proton-detected natural abundance  $^{13}\text{C}$  heteronuclear NMR spectroscopy. *J. Am. Chem. Soc.* **113**, 4371-4380 (1991).
16. Kay, L.E., Keifer, P., & Saarinen, T. Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity. *J. Am. Chem. Soc.* **114**, 10663-10665 (1992).
17. Schleucher, J., Schwendinger, M., Sattler, M., Schmidt, P., Schedletzky, O., Glaser, S.J., Sørensen, O.W., & Griesinger, C. A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J. Biomol. NMR* **4**, 301-306 (1994).

18. Davis,L.A., Keller,J., Laue,E.D., & Moskau,D. Experiments for recording pure-absorbtion heteronuclear correlation spectra using pulsed field gradients. *J. Magn. Reson.* **98**, 207-216 (1992).
19. Vuister,G.W. & Bax,A. Quantitative J Correlation: A New Approach for Measuring Homonuclear Three-Bond  $J(H^N H^\alpha)$  Coupling Constants in  $^{15}N$  Enriched Proteins. *J. Am. Chem. Soc.* **115**, 7772-7777 (1993).
20. Vuister,G.W. & Bax,A. Measurement of four-bond HN-H alpha J-couplings in staphylococcal nuclease. *J. Biomol. NMR* **4**, 193-200 (1994).
21. Karplus,M. Vicinal proton coupling in nuclear magnetic resonance. *J. Am. Chem. Soc.* **85**, 2870-2871 (1963).
22. Bertini,I., Calderone,V., Cosenza,M., Fragai,M., Lee,Y.-M., Luchinat,C., Mangani,S., Terni,B., & Turano,P. Conformational variability of MMPs: beyond a single 3D structure. *Proc. Natl. Acad. Sci. USA* **102**, 5334-5339 (2005).
23. Ruckert,M. & Otting,G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* **122**, 7793-7797 (2000).
24. Otting,G., Ruckert,M., Levitt,M.H., & Moshref,A. NMR experiments for the sign determination of homonuclear scalar and residual dipolar couplings. *J. Biomol. NMR* **16**, 343-346 (2000).
25. Schwalbe,H., Grimshaw,S.B., Spencer,A., Buck,M., Boyd,J., Dobson,C.M., Redfield,C., & Smith,L.J. A refined solution structure of hen lysozyme determined using residual dipolar coupling data. *Protein Sci* **10**, 677-688 (2001).
26. Gochin,M. & Roder,H. Protein Structure Refinement based on Paramagnetic NMR shifts. Applications to Wild-Type and mutants forms of Cytochrome c. *Protein Sci.* **4**, 296-305 (1995).
27. Zhu,G., Xia,Y., Nicholson,L.K., & Sze,K.H. Protein dynamics measurements by TROSY-based NMR experiments. *J Magn Reson* **143**, 423-426 (2000).
28. Kay,L.E., Nicholson,L.K., Delaglio,F., Bax,A., & Torchia,D.A. Pulse sequences for removal of the effects of cross correlation between dipolar and chemical-shift anisotropy relaxation mechanisms on the measurement of heteronuclear  $T_1$  and  $T_2$  values in proteins. *J. Magn. Reson.* **97**, 359-375 (1992).
29. Peng,J.W. & Wagner,G. Investigation of protein motions via relaxation measurements. *Methods Enzymol.* **239**, 563-596 (1994).
30. Pervushin,K., Riek,R., Wider,G., & Wüthrich,K. Attenuated  $T_2$  relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* **94**, 12366-12371 (1997).
31. Zeng,Y., Caignan,G.A., Bunce,R.A., Rodriguez,J.C., Wilks,A., & Rivera,M. Azide-inhibited bacterial heme oxygenases exhibit an  $S = 3/2$   $(dxz,dyz)^3(dxz)^1(dz^2)^1$  spin state: mechanistic implications for heme oxidation. *J. Am. Chem. Soc.* **127**, 9794-9807 (2005).
32. Schanda,P. & Brutscher,B. Very Fast Two-Dimensional NMR Spectroscopy for Real-Time Investigation of Dynamic Events in Proteins on the Time Scale of Seconds. *J. Am. Chem. Soc.* **127**, 8014-8015 (2005).
33. Taddese,S., Weiss,A.S., Neubert,R.H., & Schmelzer,C.E. Mapping of macrophage elastase cleavage sites in insoluble human skin elastin. *Matrix Biol.* **27**, 420-428 (2008).
34. Taddese,S., Weiss,A.S., Jahreis,G., Neubert,R.H., & Schmelzer,C.E. In vitro degradation of human tropoelastin by MMP-12 and the generation of matrikines from domain 24. *Matrix Biol.* **28**, 84-91 (2009).
35. Fields,G.B. Induction of protein-like molecular architecture by self-assembly processes. *Bioorg. Med. Chem.* **7**, 75-81 (1999).
36. Bax,A. & Davis,D.G. MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **65**, 355-360 (1985).

37. Keller, R. L. J. *The Computer Aided Resonance Assignment Tutorial*. [1.3]. 2004. CANTINA Verlag. Ref Type: Computer Program
38. Shen, Y., & Bax, A. Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J. Biomol. NMR* **38**, 289-302 (2007).
39. Meiler, J. PROSHIFT: protein chemical shift prediction using artificial neural networks. *J. Biomol. NMR* **26**, 25-37 (2003).
40. de la Torre, J.G., Huertas, M.L., & Carrasco, B. HYDRONMR: Prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* **147**, 138-146 (2000).
41. Iyer, S., Visse, R., Nagase, H., & Acharya, K.R. Crystal structure of an active form of human MMP-1. *Journal of Molecular Biology* **362**, 78-88 (2006).
42. Güntert, P., Braun, W., & Wüthrich, K. Efficient computation of three-dimensional protein structures in solution from Nuclear Magnetic Resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **217**, 517-530 (1991).
43. Wishart, D.S., Sykes, B.D., & Richards, F.M. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**, 1647-1651 (1992).
44. Shen, Y., Delaglio, F., Cornilescu, G., & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213-223 (2009).
45. Güntert, P., Mumenthaler, C., & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283-298 (1997).
46. Koradi, R., Billeter, M., & Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structure. *J. Mol. Graphics* **14**, 51-55 (1996).
47. Bertini, I., Luchinat, C., & Parigi, G. Paramagnetic constraints: an aid for quick solution structure determination of paramagnetic metalloproteins. *Concepts Magn. Reson.* **14**, 259-286 (2002).
48. Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R., & Thornton, J.M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477-486 (1996).
49. Ramachandran, G.N., Ramakrishnan, C., & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95-99 (1963).
50. Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Wang, B., Pearlman, D. A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C. E., Caldwell, J. W., Ross, W. S., and Kollman, P. A. AMBER 8. [8.0]. 2004. San Francisco, CA, University of California. Ref Type: Computer Program
51. Dominguez, C., Boelens, R., & Bonvin, A.M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731-1737 (2003).
52. Lee, B., & Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **55**, 379-400 (1971).
53. Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Pieper, U., & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* **Chapter 2**, Unit (2007).



## **3. Results**

## Evidence of Reciprocal Reorientation of the Catalytic and Hemopexin-Like Domains of Full-Length MMP-12

Ivano Bertini,<sup>\*,†,‡</sup> Vito Calderone,<sup>†</sup> Marco Fragai,<sup>†,§</sup> Rahul Jaiswal,<sup>†</sup>  
Claudio Luchinat,<sup>†,§</sup> Maxime Melikian,<sup>†</sup> Efstratios Mylonas,<sup>||</sup> and Dmitri I. Svergun<sup>||,⊥</sup>

Magnetic Resonance Center (CERM), University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy, Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy, Department of Agricultural Biotechnology, University of Florence, Via Maragliano, 75–77, 50144 Florence, Italy, European Molecular Biology Laboratory, Hamburg Outstation, Notkestrasse 85, 22603 Hamburg, Germany, and Institute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, 117333 Moscow, Russia

Received November 20, 2007; E-mail: ivanobertini@cerm.unifi.it

**Abstract:** The proteolytic activity of matrix metalloproteinases toward extracellular matrix components (ECM), cytokines, chemokines, and membrane receptors is crucial for several homeostatic and pathological processes. Active MMPs are a family of single-chain enzymes (23 family members in the human genome), most of which constituted by a catalytic domain and by a hemopexin-like domain connected by a linker. The X-ray structures of MMP-1 and MMP-2 suggest a conserved and well-defined spatial relationship between the two domains. Here we present structural data for MMP-12, suitably stabilized against self-hydrolysis, both in solution (NMR and SAXS) and in the solid state (X-ray), showing that the hemopexin-like and the catalytic domains experience conformational freedom with respect to each other on a time scale shorter than  $10^{-8}$  s. Hints on the probable conformations are also obtained. This experimental finding opens new perspectives for the often hypothesized active role of the hemopexin-like domain in the enzymatic activity of MMPs.

### Introduction

Matrix metalloproteinases (MMP) are an important family of 23 proteins which are involved in a number of extracellular processes<sup>1–3</sup> including the degradation of the extracellular matrix.<sup>4</sup> The latter is constituted by structural proteins such as collagen and elastin, by proteoglycans, and by adhesive proteins such as fibronectin, laminin, and tenascin.<sup>5</sup> MMPs are single-chain enzymes secreted by cells as inactive proenzymes. The active form is liberated outside the cell by the cleavage of the prodomain by other proteases, including MMPs themselves,<sup>6,7</sup>

thereby implying a complex regulation mechanism which also involves other proteins such as tissue inhibitors of MMPs (called TIMPs).<sup>8,9</sup>

All active MMPs but MMP-7 are constituted by two domains, a catalytic (CAT) and a hemopexin-like (HPX) domain. The CAT and HPX domains are connected by a linker whose length varies from 14 to 68 AA.<sup>10,11</sup> For many MMPs the linker is relatively short (14–23 AA) whereas for MMP-9 and MMP-15, at the other extreme, the intervening residues between the CAT and HPX domains (68 and 63 AA, respectively) constitute a further, highly glycosylated, domain termed OG domain.<sup>12,13</sup> The CAT domain alone bears full proteolytic activity toward a range of peptides and proteins.<sup>14–17</sup> However, efficient proteolysis of, for instance, triple helical collagen requires the full-

<sup>†</sup> CERM, University of Florence.

<sup>‡</sup> Department of Chemistry, University of Florence.

<sup>§</sup> Department of Agricultural Biotechnology, University of Florence.

<sup>||</sup> Hamburg Outstation.

<sup>⊥</sup> Russian Academy of Sciences.

- (1) Parks, W. C.; Wilson, C. L.; Lopez-Boado, Y. S. *Nat. Rev. Immunol.* **2004**, *4*, 617–629.
- (2) D'Alessio, S.; Fibbi, G.; Cinelli, M.; Guiducci, S.; Del Rosso, A.; Margheri, F.; Serrati, S.; Pucci, M.; Kahaleh, B.; Fan, P. S.; Annunziato, F.; Cosmi, L.; Liotta, F.; Matucci-Cerinic, M.; Del Rosso, M. *Arthritis Rheum.* **2004**, *50*, 3275–3285.
- (3) Boire, A.; Covic, L.; Agarwal, A.; Jacques, S.; Sherif, S.; Kuliopulos, A. *Cell* **2005**, *120*, 303–313.
- (4) Page-McCaw, A.; Ewald, A. J.; Werb, Z. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 221–233.
- (5) Aumailley, M.; Gayraud, B. *J. Mol. Med.* **1998**, *76*, 253–265.
- (6) Knauper, V.; Will, H.; Lopez-Otin, C.; Smith, B.; Atkinson, S. J.; Stanton, H.; Hembry, R. M.; Murphy, G. *J. Biol. Chem.* **1996**, *271*, 17124–17131.
- (7) Morrison, C. J.; Overall, C. M. *J. Biol. Chem.* **2006**, *281*, 26528–26539.

- (8) Ra, H. J.; Parks, W. C. *Matrix Biol.* **2007**, *26*, 587–596.
- (9) Overall, C. M.; Tam, E.; McQuibban, G. A.; Morrison, C.; Wallon, U. M.; Bigg, H. F.; King, A. E.; Roberts, C. R. *J. Biol. Chem.* **2000**, *275*, 39497–39506.
- (10) Maskos, K.; Bode, W. *Mol. Biotechnol.* **2003**, *25*, 241–266.
- (11) Andreini, C.; Banci, L.; Bertini, I.; Luchinat, C.; Rosato, A. *J. Proteome Res.* **2004**, *3*, 21–31.
- (12) Van den Steen, P. E.; Grillet, B.; Opendakker, G. *Glycobiol. Med.* **2005**, *564*, 45–55.
- (13) Opendakker, G.; Van den Steen, P. E.; Van Damme, J. *Trends Immunol.* **2001**, *22*, 571–579.
- (14) Gronski, T. J.; Martin, R. L.; Kobayashi, D. K.; Walsh, B. C.; Holman, M. C.; Huber, M.; VanWart, H.; Shapiro, S. D. *J. Biol. Chem.* **1997**, *272*, 12189–12194.
- (15) Mecham, R. P.; Broekelmann, T. J.; Fliszar, C. J.; Shapiro, S. D.; Welgus, H. G.; Senior, R. M. *J. Biol. Chem.* **1997**, *272*, 18071–18076.
- (16) Murphy, G.; Allan, J. A.; Willenbrock, F.; Cockett, M. I.; O'Connell, J. P.; Docherty, A. J. *J. Biol. Chem.* **1992**, *267*, 9612–9618.

length active protein.<sup>18,19</sup> For this reason, it is often hypothesized that the HPX domain helps the local unwinding of the triple helix, in such a way that a single peptide strand can be accommodated in the active site of the CAT domain and cleaved.<sup>20,21</sup> It has been also hypothesized that a relative mobility of the HPX domain is necessary for this function.<sup>22–25</sup> This seems to be the case for MMP-9, for which small-angle X-ray scattering (SAXS) and atomic force microscopy (AFM) experiments indicate that the OG domain is able to lose its globular shape and transiently assume elongated structures, thereby allowing relative motion of the CAT and HPX domains.<sup>26</sup> On the other hand, X-ray structures of the full-length proenzyme forms of other MMPs lacking the OG domain display a well-defined structural relationship between the CAT and HPX domains, and this relationship is the same in the two different pro-MMPs studied (i.e., pro-MMP-1<sup>25</sup> (14 AA linker) and pro-MMP-2<sup>27</sup> (20 AA linker)).<sup>25</sup> The structures of active MMP-1,<sup>28</sup> and of its porcine ortholog,<sup>29</sup> are also compact and show a slightly different orientation of the HPX domain with respect to pro-MMP. This difference, although small, has been highlighted as evidence of the potential ability of the HPX domain to move with respect to the CAT domain.<sup>25,28</sup>

Here we have addressed the general problem of the relative conformational freedom of the two domains of MMPs in solution by NMR. NMR in solution is a powerful tool to

investigate internal mobility of biomolecules<sup>30–42</sup> and can also provide precious information on interprotein and interdomain mobility.<sup>43–49</sup>

We selected MMP-12 (16 AA linker), for which an extended NMR assignment of the CAT domain is available, as well as high-resolution solid state and NMR structures. In this work, we have assigned the NMR signals of the HPX domain, solved its solution structure, and assigned the full-length protein. We have then obtained relaxation data ( $R_1$ ,  $R_2$ , NOE) for the full-length protein and compared it with the same data for its isolated CAT and HPX domains. These data show that the two domains are not held rigidly to one another but must undergo independent motions. Residual dipolar couplings (RDC) on the full-length protein in the presence of an external orienting device were also obtained and found inconsistent with a rigid conformation of the protein.

Repeated attempts to crystallize full-length MMP-12 for X-ray diffraction finally yielded crystals of modest quality, which diffracted to 3 Å resolution. This low resolution was, however, largely sufficient to establish that the structure is less compact and the relative orientation of the two domains totally different, with respect to the four X-ray structures of MMP-1 and MMP-2 already described. Small-angle X-ray scattering (SAXS) data are consistent with the determined structure in the crystal representing the major conformation in solution but also point to a conformational mobility of the domains in MMP-12.<sup>50</sup>

## Materials and Methods

**Protein Expression.** The cDNA encoding the G106-C470 sequence of full-length MMP-12 was generated by a polymerase chain reaction (PCR) from an IMAGE consortium clone using two synthetic oligonucleotides as primers. The cDNA obtained was cloned into the pET21a (Novagen) using the restriction enzymes *Nde I* and *Xho I* (New England BioLabs). One additional methionine at position 105 was present in the final expression product.

The expression vector encoding for the full-length protein (FL-MMP-12) was transformed into competent *Escherichia coli* BL21 (DE3) Codon Plus strain, and colonies were selected for Ampicillin and chloramphenicol resistance. Bacteria were grown in LB medium containing 34 µg/mL chloramphenicol and 50 µg/mL ampicillin in a shaker flask at 37 °C. Protein expression was induced with 0.5 mM IPTG at an OD<sub>600</sub> = 0.6, and cell growth was continued for a further 5 h. For expression of <sup>15</sup>N and <sup>13</sup>C-enriched FL-MMP-12, the bacteria were grown in minimal medium containing <sup>15</sup>N enriched (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and <sup>13</sup>C enriched glucose (Cambridge Isotope Laboratories). Cells were harvested by centrifugation and resuspended in a buffer containing 25% sucrose, 50 mM Tris-HCl (pH 8), 0.1 M NaCl, 0.2 M EDTA, 1 mM DTT. Five to ten milligrams of lysozyme were added to the resulting suspension and stirred for

- (17) Overall, C. M.; McQuibban, G. A.; Clark-Lewis, I. *Biol. Chem.* **2002**, *383*, 1059–1066.
- (18) Clark, I. E.; Cawston, T. E. *Biochem. J.* **1989**, *263*, 201–206.
- (19) Otl, J.; Gabriel, D.; Murphy, G.; Knauper, V.; Tominaga, Y.; Nagase, H.; Kroger, M.; Tschesche, H.; Bode, W.; Moroder, L. *Chem. Biol.* **2000**, *7*, 119–132.
- (20) Chung, L. D.; Dinakarandian, D.; Yoshida, N.; Lauer-Fields, J. L.; Fields, G. B.; Visse, R.; Nagase, H. *EMBO J.* **2004**, *23*, 3020–3030.
- (21) Tam, E. M.; Moore, T. R.; Butler, G. S.; Overall, C. M. *J. Biol. Chem.* **2004**, *279*, 43336–43344.
- (22) Gomis-Ruth, F. X.; Gohlke, U.; Betz, M.; Knauper, V.; Murphy, G.; Lopez-Otin, C.; Bode, W. *J. Mol. Biol.* **1996**, *264*, 556–566.
- (23) de Souza, S. J.; Pereira, H. M.; Jacchieri, S.; Brentani, R. R. *FASEB J.* **1996**, *10*, 927–930.
- (24) Overall, C. M. *Mol. Biotechnol.* **2002**, *22*, 51–86.
- (25) Jozic, D.; Bourenkov, G.; Lim, N. H.; Visse, R.; Nagase, H.; Bode, W.; Maskos, K. *J. Biol. Chem.* **2005**, *280*, 9578–9585.
- (26) Rosenblum, G.; Van den Steen, P. E.; Cohen, S. R.; Grossmann, J. G.; Frenkel, J.; Sertchook, R.; Slack, N.; Strange, R. W.; Opendakker, G.; Sagi, I. *Structure* **2007**, *15*, 1227–1236.
- (27) Morgunova, E.; Tuuttila, A.; Bergmann, U.; Isupov, M.; Lindqvist, Y.; Schneider, G.; Tryggvason, K. *Science* **1999**, *284*, 1667–1670.
- (28) Iyer, S.; Visse, R.; Nagase, H.; Acharya, K. R. *J. Mol. Biol.* **2006**, *362*, 78–88.
- (29) Li, J.; Brick, P.; Ohare, M. C.; Skarzynski, T.; Lloyd, L. F.; Curry, V. A.; Clark, I. M.; Bigg, H. F.; Hazleman, B. L.; Cawston, T. E.; Blow, D. M. *Structure* **1995**, *3*, 541–549.
- (30) Powers, R.; Clore, G. M.; Stahl, S. J.; Wingfield, P. T.; Gronenborn, A. M. *Biochemistry* **1992**, *31*, 9150–9157.
- (31) Szyperski, T.; Luginbuhl, P.; Otting, G.; Güntert, P.; Wüthrich, K. *J. Biomol. NMR* **1993**, *3*, 151–164.
- (32) Fischer, M. W. F.; Zeng, L.; Majumdar, A.; Zuiderweg, E. R. P. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8016–8019.
- (33) Ishima, R.; Torchia, D. A. *Nat. Struct. Biol.* **2000**, *7*, 740–743.
- (34) Pfeiffer, S.; Fushman, D.; Cowburn, D. *J. Am. Chem. Soc.* **2001**, *123*, 3021–3026.
- (35) Tolman, J. R.; Al-Hashimi, H. M.; Kay, L. E.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123*, 1416–1424.
- (36) Mulder, F. A. A.; Mittermaier, A.; Hon, B.; Dahlquist, F. W.; Kay, L. E. *Nat. Struct. Biol.* **2001**, *8*, 932–935.
- (37) Eisenmesser, E. Z.; Bosco, D. A.; Akke, M.; Kern, D. *Science* **2002**, *295*, 1520–1523.
- (38) Peti, W.; Meiler, J.; Brüsweiler, R.; Griesinger, C. *J. Am. Chem. Soc.* **2002**, *124*, 5822–5833.
- (39) Bruschweiler, R. *Curr. Opin. Struct. Biol.* **2003**, *13*, 175–183.
- (40) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128–132.
- (41) Mittermaier, A.; Kay, L. E. *Science* **2006**, *312*, 224–228.

- (42) Ferrage, F.; Pelupessy, P.; Cowburn, D.; Bodenhausen, G. *J. Am. Chem. Soc.* **2006**, *128*, 11072–11078.
- (43) Barbato, G.; Ikura, M.; Kay, L. E.; Pastor, R. W.; Bax, A. *Biochemistry* **1992**, *31*, 5269–5278.
- (44) Fischer, M. W.; Losonczi, J. A.; Weaver, J. L.; Prestegard, J. H. *Biochemistry* **1999**, *38*, 9013–9022.
- (45) Baber, J. L.; Szabo, A.; Tjandra, N. *J. Am. Chem. Soc.* **2001**, *123*, 3953–3959.
- (46) Jain, N. U.; Wyckoff, T. J. O.; Raetz, C. R. H.; Prestegard, J. H. *J. Mol. Biol.* **2004**, *343*, 1379–1389.
- (47) Volkov, A. N.; Worrall, J. A. R.; Holtzman, E.; Ubbink, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18945–18950.
- (48) Ryabov, Y. E.; Fushman, D. *Magn. Reson. Chem.* **2006**, *44*, S143–S151.
- (49) Ryabov, Y. E.; Fushman, D. *J. Am. Chem. Soc.* **2007**, *129*, 3315–3327.
- (50) Bertini, I.; Calderone, V.; Cosenza, M.; Fragai, M.; Lee, Y.-M.; Luchinat, C.; Mangani, S.; Terni, B.; Turano, P. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5334–5339.

15–20 min at 4 °C. A buffer containing 2% Triton, 50 mM Tris-HCl (pH 8), 0.1 M NaCl, 0.2 M EDTA, and 1 mM DTT was added, and the suspension was sonicated (7–8 30 s cycles) and centrifuged at 40 000 rpm for 20 min at 4 °C. The pellets were resuspended in 6 M Urea, 20 mM Tris-HCl (pH 8) and centrifuged. The resulting inclusion bodies were solubilized in 6 M Gdn-HCl, 20 mM Tris-HCl (pH 8), 10 mM DTT, and 20 mM cystamine and stirred overnight at 4 °C. The insoluble material was removed by centrifuging at 9000 rpm for 10 min. Protein molecular weight and purity was checked on 15% gel by SDS-PAGE. The F171D, E219A mutant of FL-MMP-12 was produced using the quick-change site-directed mutagenesis kit (Qiagen), and the expression and purification of the protein and of its <sup>15</sup>N- and <sup>13</sup>C-<sup>15</sup>N-enriched versions completed using the same procedure described above.

Refolding was carried out by a serial dilution method.<sup>51</sup> The protein was diluted to a final concentration of 0.13 mg/mL in 6 M Gdn-HCl, 20 mM Tris-HCl (pH 8), 1 mM DTT, and 0.05% Brij-35, stirred at 4 °C for 30–60 min, and dialyzed overnight against a buffer containing 20 mM Tris (pH 7.2), 10 mM CaCl<sub>2</sub>, 0.1 mM ZnCl<sub>2</sub>, 0.15 M NaCl, 5 mM β-mercaptoethanol, 1 mM 2-hydroxyethyl disulfide, and 0.05% Brij-35. The refolded protein was then purified using a Sepharose column (Amersham) with a buffer containing 20 mM Tris (pH 7.2), 10 mM CaCl<sub>2</sub>, 0.3 M NaCl and 0.1 M acetohydroxamic acid (AHA). The protein solution was concentrated up to 5 mL and purified by size exclusion chromatography using a High Load 16/60 Superdex 75 (Amersham Biosciences) and eluted with 20 mM Tris pH 7.2, 10 mM CaCl<sub>2</sub>, 0.3 M NaCl, and 0.2 M AHA. The eluted fractions were checked for purity on 15% gel by SDS-PAGE, and those containing the FL-MMP-12 protein were pooled and concentrated.

Samples of cadmium(II) substituted FL-MMP-12 protein were prepared by exhaustive dialysis against a buffer containing 20 mM Tris pH 7.2, 10 mM CaCl<sub>2</sub>, 0.3 M NaCl, 0.2 M AHA, and 0.3 mM of CdCl<sub>2</sub>.<sup>52</sup> Equimolar concentrations of *N*-isobutyl-*N*-[4-methoxyphenylsulfonyl] glycol hydroxamic acid (NNGH) were added to the samples to further increase the protein stability.

The cDNA encoding for the HPX domain (E278-C470) was generated by polymerase chain reaction and cloned into pET21a, using *Nde I* and *Xho I* as restriction enzymes. The expression vector was then transformed into competent *E. coli* BL21 (DE3) Gold strain, and the colonies were selected for Ampicillin resistance. Protein refolding for both nonlabeled and for <sup>13</sup>C and/or <sup>15</sup>N enriched samples was carried out by using the same protocols previously described for the preparation of FL-MMP-12 samples. Samples of the zinc(II) catalytic domain (F171D mutant) were prepared as previously described.<sup>50</sup>

#### NMR Measurements and Solution Structure Calculations.

The experiments for the structure calculation and mobility measurements of the isolated HPX domain were performed on protein samples at concentrations ranging between 0.5 and 0.7 mM (pH 7.2). For FL-MMP-12, all NMR experiments were performed on samples at a concentration of 0.5 mM (pH 7.2).

All NMR experiments were performed at 298 K and acquired on Bruker AVANCE 900, AVANCE 800, AVANCE 700 and DRX 500 spectrometers. All instruments but one are equipped with triple resonance CRYO-probes. The 700 MHz spectrometer is equipped with a triple resonance (TXI) 5 mm probe with a *z*-axis pulse field gradient.

All spectra were processed with the Bruker TOPSPIN software packages and analyzed by the program CARA (Computer Aided Resonance Assignment, ETH Zürich).<sup>53</sup>

The backbone resonance assignment was obtained by the analysis of HNCA, HNCOC, HN(CA)CO, HNCACB, and CBCA(CO)NH

spectra performed at 900 MHz. The assignment of the aliphatic side-chain resonances was performed through the analysis of 3D (H)CCH-TOCSY spectra at 500 MHz, together with 3D <sup>15</sup>N- and <sup>13</sup>C-NOESY-HSQC spectra at 900 MHz. The obtained assignments are reported in Tables S1 and S2 for the full length protein and tables S3 and S4 for the hemopexin domain. <sup>3</sup>*J*<sub>H<sub>NH</sub>H<sub>α</sub></sub> coupling constants were determined through the HNHA experiment at 500 MHz. Backbone dihedral φ angles were independently derived from <sup>3</sup>*J*<sub>H<sub>NH</sub>H<sub>α</sub></sub> coupling constants through the appropriate Karplus equation.<sup>54</sup> Backbone dihedral ψ angles for residue *i*-1 were also determined from the ratio of the intensities of the d<sub>αN</sub>(*i*-1,*i*) and d<sub>αN</sub>(*i*,*i*) NOEs present on the <sup>15</sup>N(*i*) plane of residue *i* obtained from the <sup>15</sup>N-edited NOESY-HSQC spectrum.

3D <sup>15</sup>N- and <sup>13</sup>C-enriched NOESY-HSQC cross peak intensities were integrated using the integration routine implemented in CARA and converted into interatomic upper distance limits by the program CALIBA.<sup>55</sup>

The protein assignment and the mobility measurements on FL-MMP-12 were performed on the NNGH-inhibited, cadmium(II)-substituted Phe171Asp/Glu219Ala mutant, due to its high stability to the self-hydrolysis. Mobility measurements on the catalytic domain where performed on the NNGH-inhibited, zinc(II) form of the Phe171Asp mutant.<sup>50</sup>

Residual dipolar couplings have been measured on FL-MMP-12 in the presence of an external orienting medium constituted by a binary mixture of C<sub>12</sub>E<sub>5</sub> (penta-ethyleneglycol dodecyl ether, Fluka) and neat *n*-hexanol (Fluka), with a molar ratio C<sub>12</sub>E<sub>5</sub>/*n*-hexanol of 0.96 and with a C<sub>12</sub>E<sub>5</sub>/water ratio of 5 wt %.<sup>56</sup> One-bond <sup>1</sup>H–<sup>15</sup>N coupling constants were measured at 298 K and 900 MHz by using the IPAP method.<sup>57,58</sup> Two-hundred fifty-nine rdc values could be measured that ranged from –46 to +25 Hz. Of them, only those rdc values corresponding to residues experiencing neither mobility nor large rmsd (140 residues, mostly in α or β secondary structures, see Table S5, Supporting Information) have been used for structure calculations and to investigate the reciprocal mobility of the two domains.

The program DYANA<sup>59</sup> was used to calculate a family of 1600 structures of the isolated HPX domain starting from randomly generated conformers in 20 000 annealing steps. The solution structure statistics are reported in Table S6 (Supporting Information). The family was energy-minimized by iterative cycles of DYANA with the program FANTAORIENT.<sup>60</sup> The quality of the structures calculated by DYANA can be assessed by a properly defined energy function (target function) proportional to the squared deviations of the calculated constraints from the experimental ones, plus the standard covalent and nonbonded energy terms. Structure calculation statistics and the structural quality were evaluated using the program PROCHECK\_NMR.<sup>61</sup>

**R<sub>1</sub>, R<sub>2</sub>, and NOE Measurements.** The experiments for the determination of <sup>15</sup>N longitudinal and transverse relaxation rates and <sup>1</sup>H–<sup>15</sup>N NOE were recorded at 298 K and 700 MHz on <sup>15</sup>N-enriched samples. The <sup>15</sup>N longitudinal relaxation rates (*R*<sub>1</sub>) were measured using a sequence modified to remove cross correlation effects during the relaxation delay.<sup>62</sup> Inversion–recovery times

(54) Vuister, G. W.; Bax, A. *J. Am. Chem. Soc.* **1993**, *115*, 7772–7777.

(55) Guntert, P.; Braun, W.; Wüthrich, K. *J. Mol. Biol.* **1991**, *217*, 517–530.

(56) Rückert, M.; Otting, G. *J. Am. Chem. Soc.* **2000**, *122*, 7793–7797.

(57) Otting, G.; Rückert, M.; Levitt, M. H.; Moshref, A. *J. Biomol. NMR* **2000**, *16*, 343–346.

(58) Otting, M.; Delaglio, F.; Bax, A. *J. Magn. Reson.* **1998**, *131*, 373–378.

(59) Guntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol.* **1997**, *273*, 283–298.

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

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

(62) Kay, L. E.; Nicholson, L. K.; Delaglio, F.; Bax, A.; Torchia, D. A. *J. Magn. Reson.* **1992**, *97*, 359–375.

(51) Bertini, I.; Calderone, V.; Fragai, M.; Luchinat, C.; Mangani, S.; Terni, B. *Angew. Chem. Int. Ed.* **2003**, *42*, 2673–2676.

(52) Bertini, I.; Fragai, M.; Lee, Y.-M.; Luchinat, C.; Terni, B. *Angew. Chem. Int. Ed.* **2004**, *43*, 2254–2256.

(53) Keller, R. L. J. *The Computer Aided Resonance Assignment Tutorial*, 1.3; Cantina: Verlag, 2004.

ranging between 2.5 and 3000 ms, with a recycle delay of 3.5 s, were used for the experiments. The  $^{15}\text{N}$  transverse relaxation rates ( $R_2$ ) were measured using a CPMG sequence<sup>62,63</sup> with delays ranging between 8.5 and 237.4 ms for the CAT domain, between 8.5 and 203.5 ms for the HPX domain, and finally between 8.5 and 135.7 ms for the FL-MMP-12 protein with a refocusing delay of 450  $\mu\text{s}$ . The relaxation data are reported in Table S7 (Supporting Information).  $R_1$  and  $R_2$  data measured on the full length protein were found noisier and less uniform with respect to those of the single catalytic and hemopexin domains. This is related to the overlap of the signals in a so large protein and to the relative low solubility of the full length construct.

**Crystallization, Data Collection, and X-ray Structure Determination.** Crystals of the zinc(II) Phe171Asp/Glu219Ala FL-MMP-12 were obtained using the vapor diffusion technique at 20 °C from a solution containing 0.1 M Tris-HCl, 30% PEG-8000, 200 mM AHA, pH 8.0. The final protein concentration was 0.7 mM.

The data collection was carried out in-house, using a PX-Ultra copper sealed tube source (Oxford Diffraction) equipped with an Onyx CCD detector. The data set was collected at 100 K and the crystals used for data collection were cryo-cooled using a solution containing 10% ethylene glycol in the mother liquor.

The crystal diffracted to 3.0 Å resolution; it belongs to space-group C2 with one molecule in the asymmetric unit, a solvent content of about 50%, and a mosaicity of about 1.0°.

The data was processed using the program MOSFLM<sup>64</sup> and scaled using the program SCALA<sup>65</sup> with the TAILS and SECONDARY corrections on (the latter restrained with a TIE SURFACE command) to achieve an empirical absorption correction. The structure was solved using the molecular replacement technique in two following steps; in the first step, the model used to find the correct orientation of the catalytic part of the structure was 1Y93 whereas the one used for the hemopexin domain was the same domain in proMMP-1 structure (1SU3). In both steps, inhibitors, water molecules, and ions were omitted from the models.

The correct orientation and translation of the molecules within the crystallographic unit cell was determined with standard Patterson search techniques<sup>66,67</sup> as implemented in the program MOLREP.<sup>68,69</sup> The isotropic refinement was carried out using REFMAC<sup>70</sup> and default weights for the crystallographic term and the geometrical term were used.

In between the refinement cycles, the models were subjected to manual rebuilding by using XtalView.<sup>71</sup> The same program was used to model AHA molecule. Water molecules have been added by using the standard procedures within the ARP/WARP suite.<sup>72</sup> The stereochemical quality of the refined models was assessed using the program PROCHECK.<sup>73</sup> The Ramachandran plot is of average quality for such a resolution structure. Of the seven violations four (Asp303, Ser311, Lys315, and Asn363) are located in the hemopexin domain; it is interesting to compare these outliers to the equivalent residues in MMP-1 (1SU3) which has a higher resolution

(2.2 Å). Asp299 (corresponding to Asp303 in MMP-12) is an outlier as well, and in the place of Ser311 there is a proline residue (Pro307) which has of course a unique stereochemistry; furthermore in the place of Asn363 there is Gly359 in MMP-1, which again is unique from the stereochemical point of view. Therefore, three out of the four residues that are in disallowed regions in the HPX domain may actually assume peculiar positions in the 3-D structure. Table S8 (Supporting Information) reports the data collection and refinement statistics.

**SAXS Experiments and Data Analysis.** Synchrotron X-ray scattering data from solutions of the NNGH-inhibited, cadmium(II)-substituted Phe171Asp/Glu219Ala double mutant of FL-MMP-12 were collected on the X33 beamline of the EMBL (DESY, Hamburg),<sup>74</sup> using a MAR345 image plate detector. The scattering patterns were measured with a 2 min exposure time for several solute concentrations in the range from 0.8 to 8.3 mg/mL. To check for radiation damage, two 2 min exposures were compared, and no changes were detected. Using the sample-detector distance of 2.7 m, a range of momentum transfer of  $0.01 < s < 0.5 \text{ \AA}^{-1}$  was covered ( $s = 4\pi \sin(\theta)/\lambda$ , where  $2\theta$  is the scattering angle, and  $\lambda = 1.5 \text{ \AA}$  is the X-ray wavelength). The data were processed using standard procedures and extrapolated to infinite dilution using the program PRIMUS.<sup>75</sup> The forward scattering,  $I(0)$ , and the radius of gyration,  $R_g$ , were evaluated using the Guinier approximation,<sup>76</sup> assuming that at very small angles ( $s < 1.3/R_g$ ) the intensity is represented as  $I(s) = I(0) \exp(-s^2 R_g^2/3)$ . The values of  $I(0)$  and  $R_g$ , as well as the maximum dimension,  $D_{\text{max}}$ , and the interatomic distance distribution functions,  $(p(r))$ , were also computed using the program GNOM.<sup>77</sup> The molecular mass of FL-MMP-12 was evaluated by comparison of the forward scattering with that for a reference solution of bovine serum albumin (66 kDa).

The scattering from the high resolution models was computed using the program CRY SOL.<sup>78</sup> Given the atomic coordinates, the program either predicts the theoretical scattering pattern or fits the experimental intensity by adjusting the excluded volume of the particle and the contrast of the hydration layer to minimize the discrepancy

$$x^2 = \frac{1}{N-1} \sum_i \left[ \frac{I_{\text{exp}}(s_i) - cI_{\text{calc}}(s_i)}{\sigma(s_i)} \right]^2 \quad (1)$$

where  $N$  is the number of experimental points,  $c$  is a scaling factor,  $I_{\text{exp}}(s_i)$  and  $I_{\text{calc}}(s_i)$  are the experimental and calculated intensities, respectively, and  $\sigma(s_i)$  is the experimental error at the momentum transfer  $s_i$ .

To assess the conformational variability of MMP-12, an ensemble optimization method (EOM) was used,<sup>79</sup> allowing for coexistence of multiple conformations in solution. About 5000 randomized models of MMP-12 differing by the conformation of the interdomain linker were generated using the program DYANA starting from randomly generated conformers of the full-length protein where only the dihedral angles of the linker region were left free to vary. These models formed a pool of possible structures, for which the scattering patterns were computed by CRY SOL. The EOM program employs a genetic algorithm to select from the pool a small number (usually about 20) of representative structures such that the average scattering from the selected ensemble fits the experimental data. Multiple runs of EOM were performed and the results were averaged to provide quantitative information about

- (63) Peng, J. W.; Wagner, G. *Methods Enzymol.* **1994**, *239*, 563–596.  
 (64) Leslie, A. G. W. In *Molecular data processing*; Moras, D., Podjarny, A. D., Thierry, J.-C., Eds.; Oxford University Press: Oxford, 1991; pp 50–61.  
 (65) Evans, P. R. *Joint CCP4 and ESF-EACBM Newsletter* **1997**, *33*, 22–24.  
 (66) Rossmann, M. G.; Blow, D. M. *Acta Crystallogr.* **1962**, *15*, 24.  
 (67) Crowther, R. A. *The Molecular Replacement Method*; Rossmann, M. G., Ed.; Gordon & Breach: New York, 1972.  
 (68) Vagin, A.; Teplyakov, A. *J. Appl. Crystallogr.* **1997**, *30*, 1022–1025.  
 (69) Vagin, A.; Teplyakov, A. *Acta Crystallogr. D: Biol. Crystallogr.* **2000**, *56*, 1622–1624.  
 (70) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr.* **1997**, *D53*, 240–255.  
 (71) McRee, D. E. *J. Mol. Graphics* **1992**, *10*, 44–47.  
 (72) Lamzin, V. S. *Acta Crystallogr. D: Biol. Crystallogr.* **1993**, *49*, 129–147.  
 (73) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.

- (74) Roessle, M. W.; Klaering, R.; Ristau, U.; Robrahn, B.; Jahn, D.; Gehrman, T.; Konarev, P. V.; Round, A.; Fiedler, S.; Hermes, S.; Svergun, D. I. *J. Appl. Crystallogr.* **2007**, *40*, s190–s194.  
 (75) Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I. *J. Appl. Crystallogr.* **2003**, *36*, 1277–1282.  
 (76) Guinier, A. *Ann. Phys.* **1939**, *12*, 161–237.  
 (77) Svergun, D. I. *J. Appl. Crystallogr.* **1992**, *25*, 495–503.  
 (78) Svergun, D. I.; Barberato, C.; Koch, M. H. J. *J. Appl. Crystallogr.* **1995**, *28*, 768–773.  
 (79) Bernado, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. *J. Am. Chem. Soc.* **2007**, *129*, 5656–5664.

the flexibility of the protein in solution (in particular, about the  $R_g$  distribution in the selected ensembles).

## Results and Discussion

**Design and Production of a Stable Full-Length MMP-12.** Several MMPs, both as isolated catalytic domains (CAT) or full-length (FL) proteins, are relatively unstable *in vitro* due to self-proteolysis (while the inactive isolated hemopexin domain (HPX) is stable). In the case of MMP-12, the CAT domain can be stabilized by a Phe171Asp mutation that prevents self-proteolysis and increases the solubility. This mutation is far both from the active site and from the CAT-HPX interface (see later) and does not perturb the catalytic activity of the enzyme. Therefore, a FL construct with the Phe171Asp mutation was initially produced. The resulting protein was well folded and fully active but still sensitive to the self-proteolysis at the linker region even in the presence of the inhibitor NNGH.

To increase the stability, the catalytically relevant Glu219 was mutated to an Ala. This mutation had been shown to decrease the catalytic activity of the isolated CAT domain of MMPs by two-3 orders of magnitude,<sup>80</sup> whereas the three-dimensional structure of the domain is fully retained.<sup>81</sup> The Phe171Asp/Glu219Ala double mutant of the FL protein was stable for several days.

Finally, we substituted the catalytic zinc(II) ion, responsible for the residual activity, with a cadmium(II) ion. The resulting derivative, complexed with NNGH, resulted stable to proteolysis, showing no trace of cleaved CAT and HPX domains in the gel after three weeks of NMR measurements. The analysis of the HSQC spectra revealed that no structural alteration were caused by the replacement of zinc(II) with the cadmium(II) ion.

The NNGH-inhibited, cadmium(II)-substituted Phe171Asp/Glu219Ala double mutant of FL-MMP-12 was therefore used in all NMR experiments reported hereafter.

**Low-Resolution X-Ray Structure of FL-MMP-12 and Comparison with Existing Structural Information on FL-MMP.** Crystals diffracting at 3.0 Å were obtained for the AHA-inhibited, zinc(II) Phe171Asp/Glu219Ala double mutant of FL-MMP-12.

The X-ray structure of this construct was solved from diffraction data obtained in-house and deposited in PDB (PDB code: 3BA0). Despite the low resolution, the electron density was of generally good quality throughout the entire molecule, the only exception being residues 271–274 in the middle of the linker region (see below), which spans from Asp264 to Pro279. The overall structure of the CAT domain is very similar to the high resolution structure of the isolated CAT domain.<sup>50</sup> The mutations are both clearly visible; the active site contains both the native catalytic zinc ion and the structural zinc ion. Three calcium atoms are also present in the structure of the CAT domain. Finally, a AHA molecule bound to the zinc ion in the usual geometry<sup>50</sup> is present at the active site. The HPX domain has the expected four-blade propeller, hemopexin-like

fold previously observed in other HPX domains of MMPs.<sup>22,25,27,29,82–85</sup> The disulfide bridge between Cys282 and Cys470 is clearly present, and a calcium ion is bound in the central region of the domain.

The relative orientation of the two domains could be unambiguously determined from the 3 Å X-ray data (Figure 1A). The two domains are in contact through a relatively small surface (~165 Å<sup>2</sup>) which includes a possible salt bridge between His112 of CAT domain and the C-terminal of Cys470 in the HPX domain, and van der Waals interactions between residue 113 of CAT and residues 284 and 463 on the HPX side. The first part of the linker (residues 264, 265, 266, 270, 271, and 273) is in touch with the CAT domain (residues 112, 143, 144, 249, 259, and 263) through a surface of 902 Å<sup>2</sup> and the last part (residues 277 and 279) with the HPX domain (residues 280, 307, and 470) through a surface of 571 Å<sup>2</sup> (Figure 1A). Furthermore, the linker residue Lys266, besides being in contact with the CAT domain, forms a salt bridge with the C-terminal of HPX domain. The central part of the linker (residues 271–274) shows a rather poor electron density, but its structure can be reasonably modeled to match the arrangement of the initial and final linker residues. It is apparent that the whole molecule is held together mainly by interactions between each single domain and the linker rather than between the two domains. The presence of the intact linker was confirmed by a gel experiment on a protein solution obtained after redissolving crystals collected from the same wells of those used for X-ray.

Figure 1C shows a superposition of the X-ray structures of human proMMP-1 and proMMP-2, the only FL-proMMPs structures available to date, where the prodomain and the fibronectin domains (in case of MMP-2) have been omitted for clarity. The CAT-HPX interface is very similar in the two proteins and relatively extended (ca. 753 and 866 Å<sup>2</sup>, respectively, as opposed to ca. 165 Å<sup>2</sup> in FL-MMP-12), suggestive of a stable domain-domain interaction. Figure 1D shows a superposition of the structures of active human FL-MMP-1<sup>28</sup> and of its porcine ortholog.<sup>29</sup> Again, the two structures are very similar, and the CAT-HPX interface is again extended (ca. 735 and 747 Å<sup>2</sup>, respectively). The relative orientation of the HPX domain is slightly different in these active forms with respect to the pro-enzyme forms (cfr. Figure 1D and 1C, where the CAT domain has the same orientation), as previously noticed.<sup>28</sup>

Relative to the CAT domain, the HPX domain in FL-MMP-12 lies at about 120° with respect to its orientation in the structures of MMP-1 and MMP-2 (Figure 1E), and the two domains have a less compact arrangement (cfr. Figure 1A and 1B). The interface region is completely different, both viewed from the CAT and from the HPX domain, is sensibly smaller (ca. 165 vs 735–866 Å<sup>2</sup>), and the complementarity between the surfaces of the two domains is poor with respect to that observed in the experimental structures of FL-MMP-1 and FL-MMP-2.

**NMR Characterization of FL-MMP-12.** The NNGH-inhibited, cadmium(II)-substituted Phe171Asp/Glu219Ala double mutant of FL-MMP-12 (FL-MMP-12 hereafter) yields <sup>15</sup>N–<sup>1</sup>H HSQC spectra of surprisingly good quality for a protein of 42 000 Da (see Figure 2A) (spectra of this type were obtained on all constructs of FL-MMP-12 described before).

Figure 2B shows the <sup>15</sup>N–<sup>1</sup>H HSQC spectrum of FL-MMP-12 (in black) superimposed to those of the isolated CAT domain (in green), prepared as previously reported, and of the isolated HPX domain (in red, see below). The similarity is striking: except for a number of peaks that are not present in either the

(80) Cha, J.; Auld, D. S. *Biochemistry* **1997**, *36*, 16019–16024.

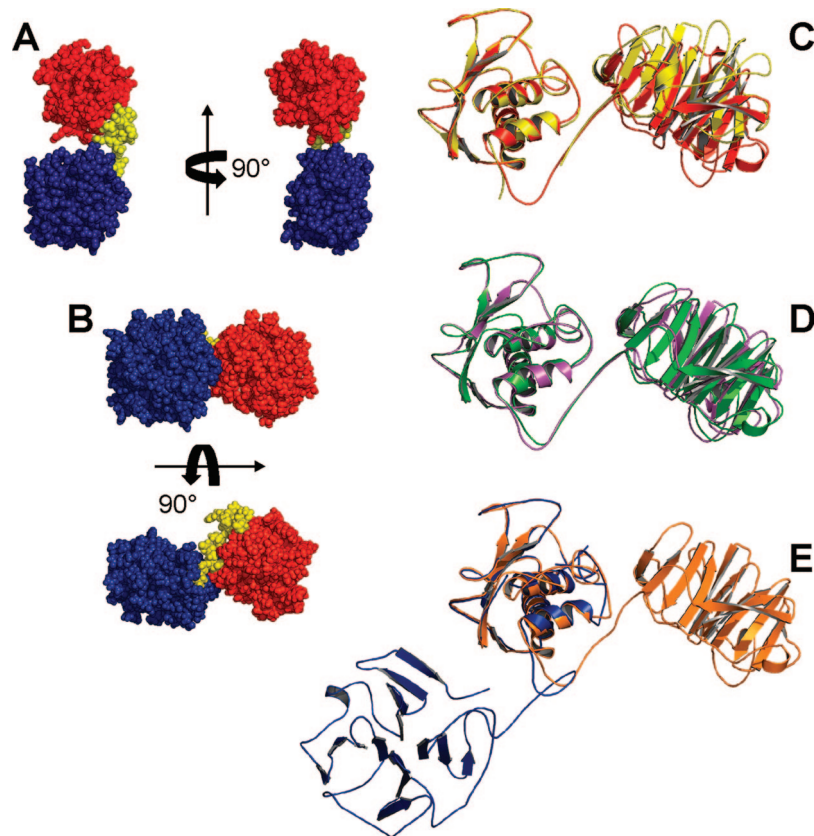
(81) Bertini, I.; Calderone, V.; Fragai, M.; Luchinat, C.; Maletta, M.; Yeo, K. J. *Angew. Chem., Int. Ed.* **2006**, *45*, 7952–7955.

(82) Morgunova, E.; Tuuttila, A.; Bergmann, U.; Tryggvason, K. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7414–7419.

(83) Libson, A. M.; Gittis, A. G.; Collier, I. E.; Marmer, B. L.; Goldberg, G. I.; Lattman, E. E. *Nat. Struct. Biol.* **1995**, *2*, 938–942.

(84) Gohlke, U.; Gomis-Ruth, F. X.; Crabbe, T.; Murphy, G.; Docherty, A. J.; Bode, W. *FEBS Lett.* **1996**, *378*, 126–130.

(85) Cha, H.; Kopetzki, E.; Huber, R.; Lanzendorfer, M.; Brandstetter, H. *J. Mol. Biol.* **2002**, *320*, 1065–1079.



**Figure 1.** Space-filling representations of the X-ray structures of FL-MMP-12 (A, present work) and FL-MMP-1 (B).<sup>28</sup> Superimposition of the X-ray structures of (C) human pro-MMP-1<sup>25</sup> (red) and pro-MMP-2<sup>27</sup> (yellow) (where the prodomain of both and the fibronectin domains of MMP-2 have been omitted for clarity), (D) human<sup>28</sup> (violet) and porcine<sup>29</sup> (green) FL-MMP-1, and (E) human FL-MMP-1<sup>28</sup> (violet) with the present FL-MMP-12 (blue).

isolated CAT and HPX domains, which must therefore belong to the linker, all other FL-MMP-12 peaks are superimposable, or nearly superimposable, to a peak of either the isolated CAT or HPX domain. Furthermore, the peaks of FL-MMP-12 were only marginally broader than the corresponding peaks in the isolated domains.

The <sup>13</sup>C, <sup>15</sup>N, and <sup>1</sup>H spectral assignment of the CAT domain was available,<sup>50</sup> whereas that of the HPX domain was performed during this research. Also the solution structure of the HPX domain, up to now unavailable for MMP-12, was experimentally solved here (see Table S6 and Figure S1, Supporting Information) and deposited in PDB (PDB code: 2JXY). Ribbon representations of the solution and crystal structures are reported in Figure S1. The relatively high rmsd between the two structures (BB = 1.88 Å, secondary structure elements = 1.16 Å) is within the indeterminacy of both the X-ray (3.0 Å resolution) and NMR structures (BB rmsd = 1.38 Å, heavy atom rmsd = 2.13 Å).

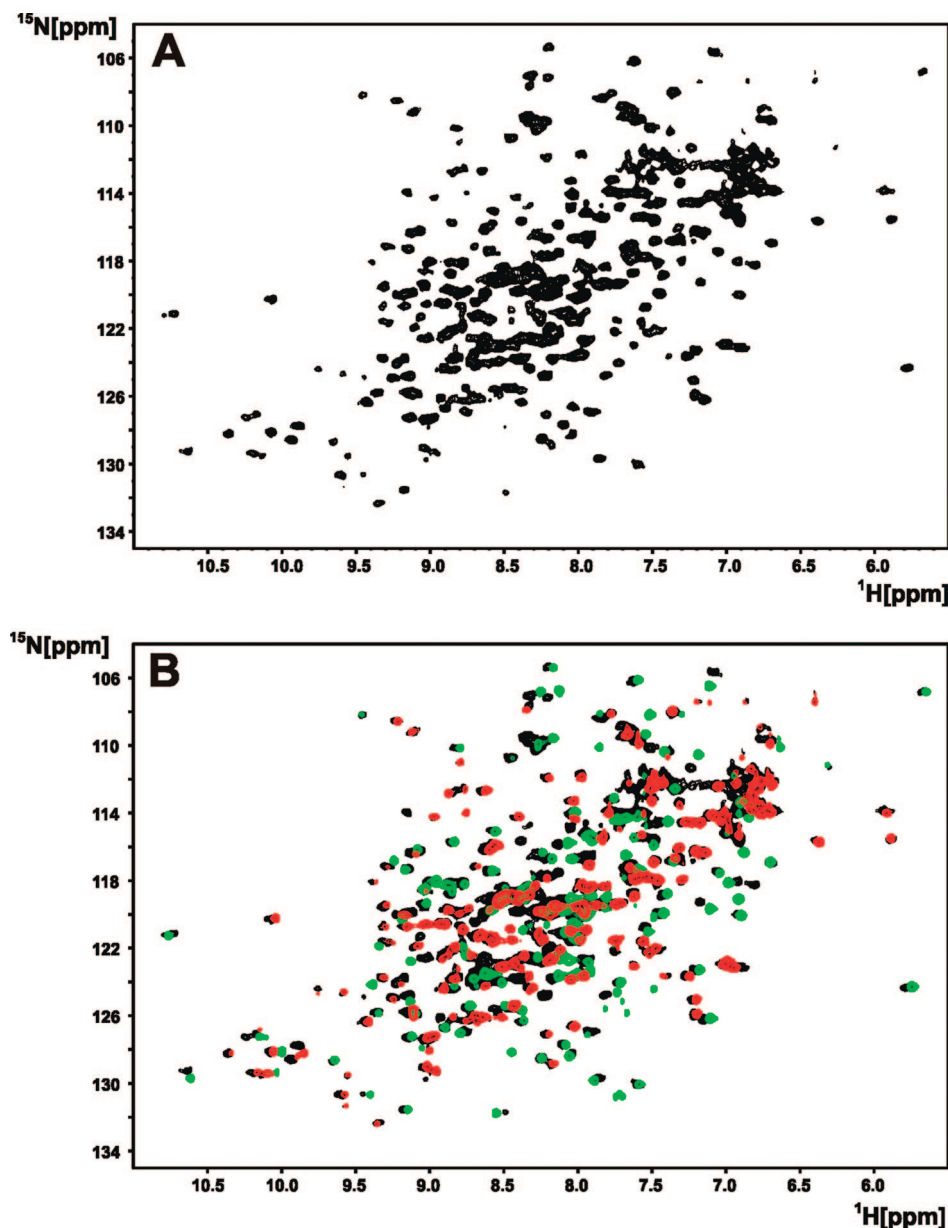
The <sup>15</sup>N–<sup>1</sup>H HSQC spectrum of FL-MMP-12 could be largely assigned by direct comparison with the isolated CAT and HPX spectra. Nevertheless, 97% of the nonlinker residues were reassigned by recording the same set of 3D spectra used for the structure of the isolated domains. Despite the crowding, the spectra were of good enough quality to assign most of the backbone atoms of the two domains as well as a large fraction of side-chain atoms (Table S1, Supporting Information). 3D NOESY spectra were also recorded. The NOESY patterns were largely the same as those observed in the isolated CAT and HPX domains, indicating that the domain structures are indeed unaltered in the FL protein, as it was already apparent from the comparison of the HSQC spectra. No interdomain NOEs could

be found. The largest difficulties in the assignment were encountered in the linker region. Although several of the linker <sup>15</sup>N–<sup>1</sup>H HSQC peaks could be identified as the extra peaks present in the FL protein spectra and not corresponding to peaks of any of the two domains, several of the necessary sequential connectivities were missing in the 3D spectra. In addition, very few NOEs could be seen for these peaks. A protonless CON experiment<sup>86</sup> allowed us to identify two of the linker prolines. Ten linker peaks could be identified, and four of them sequence-specifically assigned (Table 1).

**Relaxation Data.** The 700 MHz *R*<sub>1</sub> and *R*<sub>2</sub> data for FL-MMP-12 as well as for the isolated CAT and HPX domains are shown in Figure 3A–D. *R*<sub>1</sub> and *R*<sub>2</sub> values can be accurately estimated from the atomic coordinates of a macromolecule of known structure assuming a rigid-body hydrodynamics by using computer programs like HYDRONMR.<sup>87</sup> In this study, the X-ray structure of the CAT domain (PDB code: 1Y93),<sup>50</sup> the present solution structure of the HPX domain, and the present X-ray structure of the full length protein (Figure 1E) were used as input files in HYDRONMR. As shown in Figure 3, while for the isolated domains the experimental data nicely match the calculated values, for FL-MMP-12 sizably faster longitudinal and sizably slower transverse relaxation rates with respect to the calculated values were measured. These differences are clearly statistically significant, despite the larger errors due to the signal overlap and lower signal-to-noise ratios in the spectra

(86) Bermel, W.; Bertini, I.; Duma, L.; Emsley, L.; Felli, I. C.; Pierattelli, R.; Vasos, P. R. *Angew. Chem., Int. Ed.* **2005**, *44*, 3089–3092.

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



**Figure 2.** 900 MHz  $^{15}\text{N}$ – $^1\text{H}$  HSQC spectra of the uniformly  $^{15}\text{N}$ -labelled NNGH-inhibited, cadmium(II)-substituted Phe171Asp/Glu219Ala double mutant of FL-MMP-12 (A), and superimposed spectra of FL-MMP-12 (black), CAT-MMP-12 (green), and HPX-MMP-12 (red) (B).

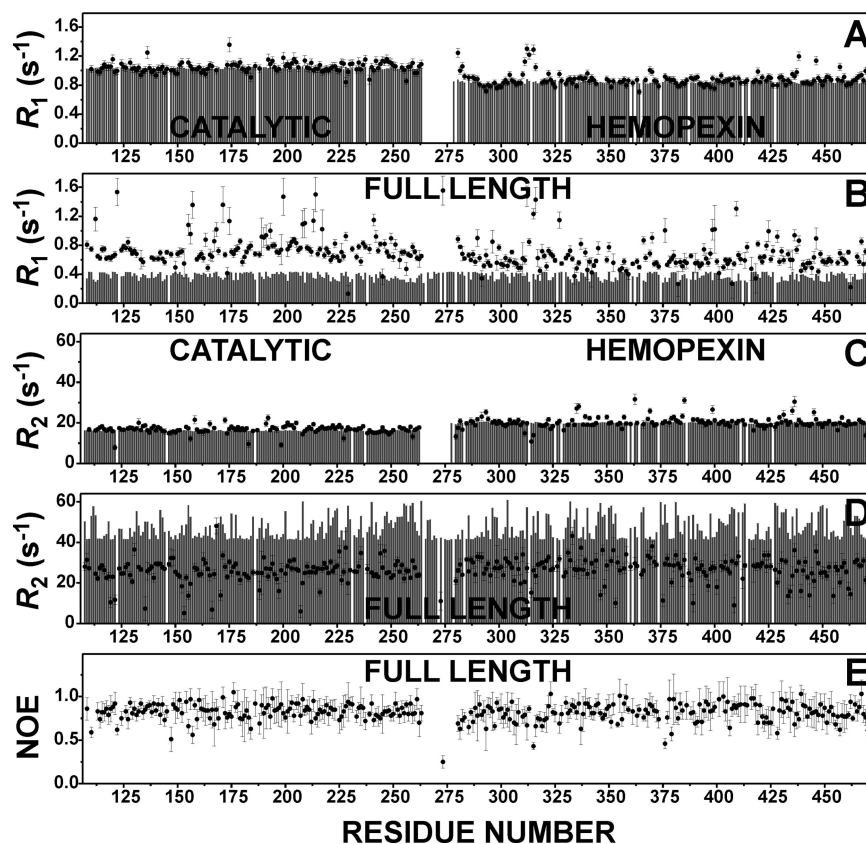
**Table 1.**  $^1\text{H}$ ,  $^{15}\text{N}$ , and  $^{13}\text{C}$  Chemical Shifts for the Identified Linker Residues of FL-MMP-12

		HN	N	CA	HA	CB	other	NOE
<b>264</b>	<b>D</b>	7.88	116.40	48.44	4.55	36.50	HB2 3.77, HB3 2.53	
<b>272</b>	<b>P</b>			60.17		29.20	CG 24.29, CD 46.96	
<b>273</b>	<b>N</b>	8.38	119.79	48.44	4.78	35.86	HB2 2.72, HB3 2.56	0.25
<b>279</b>	<b>P</b>			59.92	4.20	29.18	HB2 2.12, HB3 1.67, HG2 1.87, HG3 1.5, HD2 3.18, HD3 3.84	
<b>L1</b>		8.28	120.71	53.34	4.17	29.97	QB 1.78, CG 38.97, QG 2.87	
<b>L2</b>		8.17	109.42	57.03	4.83	39.62	HB2 3.05, HB3 2.60	
<b>L3</b>		7.89	122.58	49.62	4.18			
<b>L4</b>		8.16	121.91	53.24	4.19	30.91		0.28
<b>L5</b>		7.34	107.95	56.64				0.35
<b>L6</b>		8.37	123.69	51.65				

of FL-MMP-12. The  $R_1$  and  $R_2$  values measured on FL-MMP-12 are intermediate between the expectation from the isolated domains and a rigid full length structure. At the same time, the NOE values for FL-MMP-12 (Figure 3E) demonstrate that the single domains (catalytic and hemopexin) forming the full length protein behave as rigid bodies. Therefore, relaxation data can

be collectively taken as evidence that the full length protein does not exhibit a rigid body hydrodynamics but experience flexibility. Such flexibility must depend on the presence of a flexible linker that permits sizable reciprocal mobility of the two domains on a time scale that is faster than the reorientation time of the whole molecule.





**Figure 3.** Calculated (grey bars) and experimental (filled circles) backbone  $^{15}\text{NH}$   $R_1$  (A, B) and  $R_2$  (C, D) values for the isolated CAT and HPX domains (A, C) and for the full-length protein (B, D). Although the agreement between experimental and calculated  $R_1$  and  $R_2$  values for the isolated domains is excellent, for the full-length protein the experimental  $R_1$  values are sizably larger (B) and the  $R_2$  values sizably smaller (D) than the ones calculated for the rigid X-ray structure. Experimental NOE values for the full-length protein (E).

Indeed, three linker peaks (N273, L4, L5) display grossly altered relaxation values, and particularly  $^{15}\text{N}$ - $^1\text{H}$  NOEs that are sizably smaller than expected (0.25, 0.28, and 0.35 respectively; Table 1). This is another indication that the linker is at least partially involved in some fast conformational rearrangement, consistent with a sizable degree of reciprocal mobility of the two domains. Incidentally, similar NOE values were observed in the linker region of calmodulin, a two-domain protein known to sample an extremely large conformational space,<sup>43,45,50,88</sup> and more recently in the two-domain xylanase Cex.<sup>89</sup> All  $R_1$  and  $R_2$  data on FL-MMP-12 and on its isolated CAT and HPX domains, together with NOE data, are reported in the Supporting Information (Table S7).

**Residual Dipolar Couplings.** Residual dipolar couplings (RDC) in the presence of the external orienting device C12E5/hexanol<sup>56</sup> have been measured. They can be fitted very well to the structures of the two isolated domains (Figure 4A,B) separately, but with sensibly different orientation tensor values. The data are, instead, in striking disagreement with the solid state X-ray structure (Figure 4C). Figure 4D shows that the agreement is modest also for any of the four rigid two-domain structures that can be obtained by fitting both domains to a single orientation tensor. None of these four structures bears a resemblance with the X-ray structure in terms of relative domain

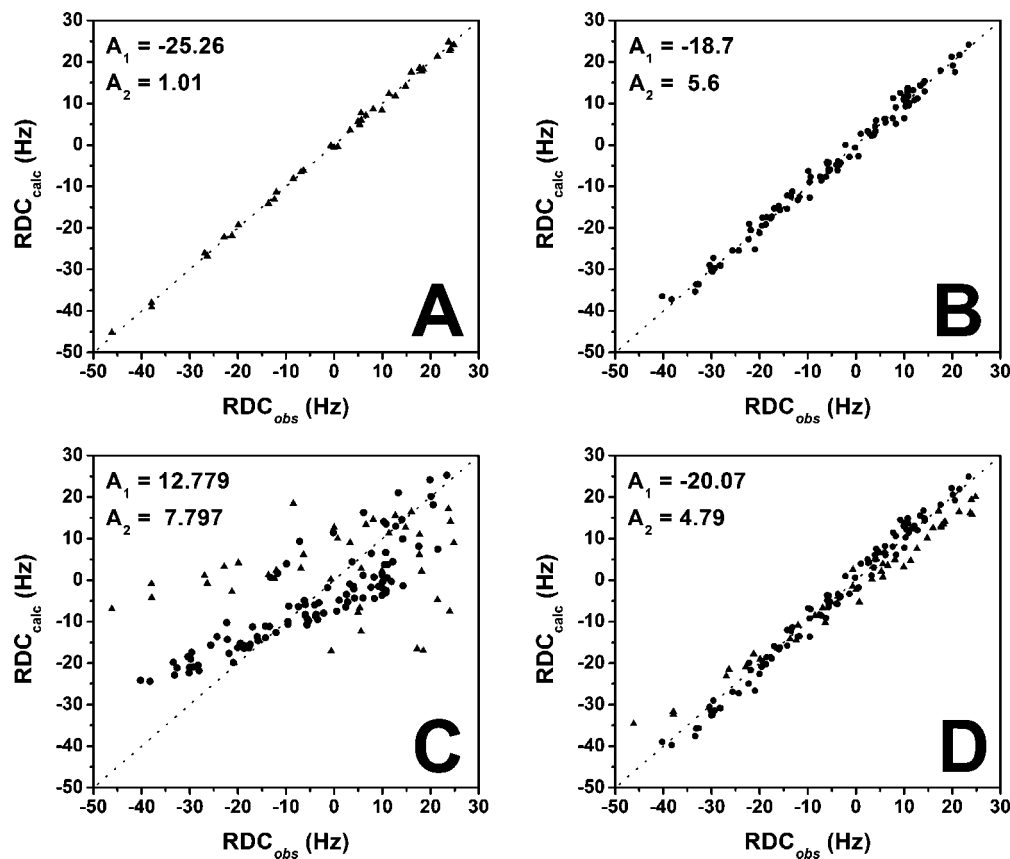
orientation (Figure S2, Supporting Information). These structures are also different from any other X-ray structure of FL-MMPs. On this basis, and on the basis of the relaxation data, these solutions are discarded.

**SAXS Experiments.** The processed X-ray scattering pattern from FL-MMP-12 presented in Figure 5 yields a molar mass estimate of  $40 \pm 4$  kDa, compatible with that calculated from the sequence (42.5 kDa), indicating that the protein is monomeric in solution. The experimental radius of gyration  $R_g$  and maximum size  $D_{\text{max}}$  are  $31 \pm 1$  Å and  $110 \pm 10$  Å, respectively. These values significantly exceed the parameters calculated from the X-ray structure of FL-MMP-1 ( $R_g = 25$  Å,  $D_{\text{max}} = 85$  Å), while they are in better agreement with those computed from the less compact X-ray structure of FL-MMP-12 determined in the present study ( $R_g = 29$  Å,  $D_{\text{max}} = 95$  Å). Moreover, the scattering pattern calculated from the FL-MMP-1 model using CRY SOL<sup>78</sup> fails to fit the experimental data (discrepancy  $\chi = 5.8$ , curve 2 in Figure 5).

The scattering curve computed from the present FL-MMP-12 structure displays a much better agreement to the experiment ( $\chi = 2.5$ , curve 3 in Figure 5), but still displays some systematic deviations. For flexible MMP-12, neither individual models, nor averaging over the random pool allowed one to fit the SAXS data. The representative ensembles selected to fit the data give information about the preferable conformations of the protein. Given the potential conformational flexibility of MMP-12 in solution suggested by the NMR data, an alternative analysis approach was applied, allowing for the coexistence of multiple protein configurations. A large number of models was generated

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

(89) Poon, D. K. Y.; Withers, S. G.; McIntosh, L. P. *J. Biol. Chem.* **2007**, *282*, 2091–2100.



**Figure 4.** Best fit vs experimental 900 MHz RDC values for the NNGH-inhibited, cadmium(II)-substituted Phe171Asp/Glu219Ala double mutant of FL-MMP-12 in a binary mixture of  $C_{12}E_5$  (penta-ethyleneglycol dodecyl ether) and neat *n*-hexanol. The separate fits for the CAT and the HPX domains and the best fit orientation tensor values are shown in A and B respectively. The poor global fit assuming the rigid X-ray structure is shown in C, and that obtained by best fitting the reciprocal orientation of the two domains to the RDC values is shown in D. In all panels, triangles refer to the CAT and filled circles refer to the HPX domain.

obtained by the linker randomizations representing possible conformations of MMP-12 in solution. None of these random models yielded a computed scattering in agreement with the experimental data. This was not unexpected, as SAXS “sees” the conformational and orientational average over the large (approx.  $10^{16}$ ) ensemble of protein molecules in the illuminated specimen volume. A simple average intensity of the generated pool also did not agree with the experiment, suggesting a nonrandom configuration of the linker in MMP-12. To assess the preferable conformations in solution, the EOM method<sup>79</sup> was used. Given a representative pool of (random) structures, the method employs a genetic algorithm to select the ensembles from this pool that best fit the experimental data, as explained in materials and methods. Several EOM runs yielded reproducible ensembles neatly fitting the experimental data with discrepancy  $\chi$  in the range 1.1–1.3, and a typical fit provided by the ensemble selected by EOM is given in Figure 5, curve 4.

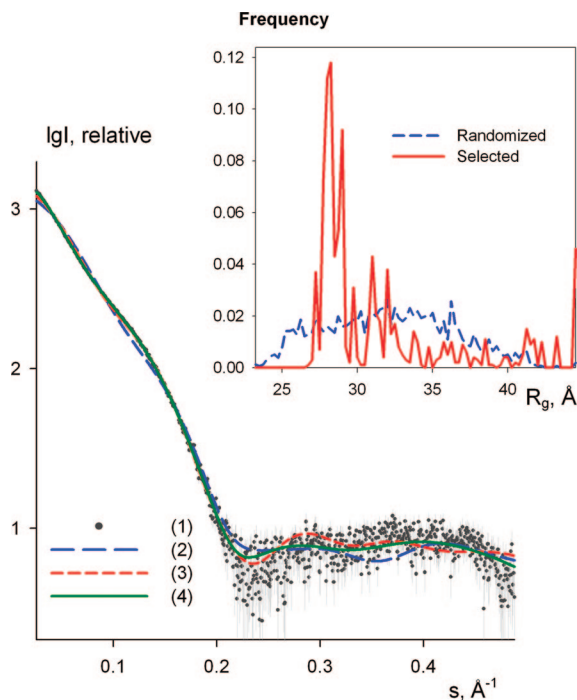
All the fits from different EOM runs are graphically indistinguishable from curve 4 in Figure 5.

The  $R_g$  distributions of the particles in the initial pool and in the selected ensembles are compared in the insert to Figure 5. The former distribution is rather broad, and covers the  $R_g$  range from about 23 to 50 Å, corresponding to extremely compact and completely extended domain configurations, respectively. In contrast, the  $R_g$  distribution of the selected ensembles displays a relatively sharp peak around  $R_g = 28$ –29 Å, including about 55% of the particles in the selected ensembles. Visual inspection of the models in the peak indicates,

not unexpectedly, that they have a shape similar to that of the MMP-12 structure in the crystal (although with varying inter-domain orientations). In contrast, not a single structure with  $R_g < 27$  Å was selected in multiple EOM reconstructions, indicating that models similar to the crystal structure of FL-MMP-1 were never present. These results indicate that the present crystal structure of FL-MMP-12 may be significantly present also in solution, but also that the protein experiences noticeable conformational flexibility, as revealed by the presence of a more or less uniform distribution of particles in the range between  $R_g = 30$  and 50 Å and of a significant spike, which was always observed for the most extended particles (Figure 5, insert). We also tried to explore the possibility of a two-state exchange situation allowing for only two conformations in the mixture. The two-state fits were however always poorer than those from twenty-state EOM populations. In particular, by fixing the first state to be the crystal structure and allowing EOM to select the second state, discrepancies not better than 1.5–1.6 could be obtained. These results suggest that MMP12 adopts a manifold of conformations in solution, in full agreement with NMR observations.

### Concluding Remarks and Biological Implications

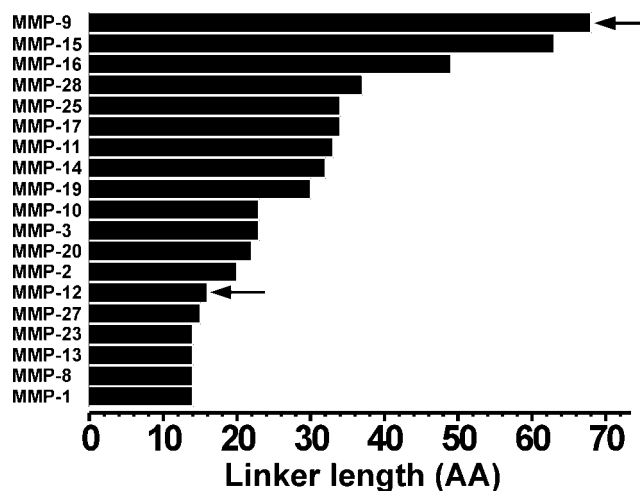
The present data demonstrate that full-length MMP-12 shows relative mobility of its catalytic and hemopexin domains. The observation of  $R_1$  and  $R_2$  values intermediate between those of the isolated domains and those expected for any rigid structure of the full-length protein is particularly striking in this respect.



**Figure 5.** Experimental X-ray scattering from the NNGH-inhibited, cadmium(II)-substituted Phe171Asp/Glu219Ala double mutant of FL-MMP-12, and scattering from the models. (1) experimental data with error bars; (2–3) computed scattering from the crystallographic models of FL-MMP-1 and FL-MMP-12, respectively; (4) a typical fit by the selected ensemble of structures. The logarithm of the scattering intensity is plotted against the momentum transfer. Inset: the frequency of the models with the given  $R_g$  in the initial pool of structures with randomized interdomain linkers (blue broken line) and in the selected ensembles (red solid line); the latter distribution is obtained by the averaging of several EOM runs. Both  $R_g$  distributions are normalized to the integral value of unity.

Indeed, even in the case of calmodulin, the two-domain protein that constitutes a paradigmatic example of large interdomain mobility,<sup>43,45,90,91</sup> the  $R_1$  and  $R_2$  values are only modestly different from what expected for a rigid structure.<sup>45,90</sup> Apparently, in the case of FL-MMP-12 the reorientation of the backbone NH vectors with respect to the magnetic field occurs on a time scale that is faster than the rotational time of the whole molecule, whereas in calmodulin it is of the same order.<sup>45,90</sup> Conversely, the amplitude of the motion is probably lower for FL-MMP-12 than it is for calmodulin, as judged from the SAXS data that suggest that the molecules spend about half of the time in a conformation that is more or less as compact as the solid state structure. A similar behavior is probably experienced by the two-domain protein xylanase Cex, whose flexibility has been recently demonstrated.<sup>89</sup>

The present data are a significant example of the synergy between NMR and SAXS techniques.<sup>92–96</sup> NMR provides evidence of conformational freedom and of the time scale, whereas SAXS provides insight into the types and variety of the sampled conformations. About half of the conformations that are likely to be experienced by FL-MMP-12 in solution are as compact as the solid state structure but with different relative orientations of the two domains, while another half are more extended, and some even highly extended. In this



**Figure 6.** Linker lengths in matrix metalloproteinases. Arrows indicate MMPs for which interdomain mobility has been demonstrated (26 and present work).

respect the X-ray structure does not provide meaningful information on the conformation of the protein in solution.

Relative mobility of the CAT and HPX domains has been recently suggested for MMP-9, where the two domains are separated by the OG domain.<sup>97</sup> In the case of MMP-9 it has been argued<sup>26</sup> that the long and flexible OG domain may mediate protein–substrate interactions. Independent domain movements might even mediate enzyme translocation on a collagen fibril.<sup>98,99</sup> Another interesting possibility is that domain flexibility can mediate the activation of the enzyme and the cleavage of the pro-domain by promoting long-range conformational transitions induced by the binding of the activator proteins.<sup>100</sup> It is possible that MMP-12 can be representative of all other MMPs where the two domains are connected by a short linker rather than by a long one (Figure 6). However, this has to be demonstrated for those cases in which the contact area between the CAT and HPX domains is much larger than in the present case. Notably, the possibility of reorienting the hemopexin with respect to the catalytic domain during catalysis has been often invoked to rationalize the fact that, for collagenases (MMP-1, MMP8, and MMP-13), the catalytic domain alone is not able to attack collagen, whereas the full-length protein does.<sup>18,19</sup>

On the contrary, noncollagenase MMPs such as MMP-2 and MMP-12 do not degrade the native triple helix collagen. The

(90) Wang, T. Z.; Frederick, K. K.; Igumenova, T. I.; Wand, A. J.; Zuiderweg, E. R. P. *J. Am. Chem. Soc.* **2005**, *127*, 828–829.  
 (91) Bertini, I.; Gupta, Y. K.; Luchinat, C.; Parigi, G.; Peana, M.; Sgheri, L.; Yuan, J. *J. Am. Chem. Soc.* **2007**, *129*, 12786–12794.

(92) Mattinen, M. L.; Paakkonen, K.; Ikonen, T.; Craven, J.; Drakenberg, T.; Serimaa, R.; Waltho, J.; Annala, A. *Biophys. J.* **2002**, *83*, 1177–1183.  
 (93) Perry, A.; Tambyrajah, W.; Grossmann, J. G.; Lian, L. Y.; Scrutton, N. S. *Biochemistry* **2004**, *43*, 3167–3182.  
 (94) Grishaev, A.; Wu, J.; Trehwella, J.; Bax, A. *J. Am. Chem. Soc.* **2005**, *127*, 16621–16628.  
 (95) Marino, M.; Zou, P. J.; Svergun, D.; Garcia, P.; Edlich, C.; Simon, B.; Wilmanns, M.; Muhle-Gol, C.; Mayans, O. *Structure* **2006**, *14*, 1437–1447.  
 (96) Tsutakawa, S. E.; Hura, G. L.; Frankel, K. A.; Cooper, P. K.; Tainer, J. A. *J. Struct. Biol.* **2007**, *158*, 214–223.  
 (97) Van den Steen, P. E.; Van Aelst, I.; Hvidberg, V.; Piccard, H.; Fiten, P.; Jacobsen, C.; Moestrup, S. K.; Fry, S.; Royle, L.; Wormald, M. R.; Wallis, R.; Rudd, P. M.; Dwek, R. A.; Opdenakker, G. *J. Biol. Chem.* **2006**, *281*, 18626–18637.  
 (98) Overall, C. M.; Butler, G. S. *Structure* **2007**, *15*, 1159–1161.  
 (99) Saffarian, S.; Collier, I. E.; Marmer, B. L.; Elson, E. L.; Goldberg, G. *Science* **2004**, *306*, 108–111.  
 (100) Rosenblum, G.; Meroueh, S.; Toth, M.; Fisher, J. F.; Fridman, R.; Mobashery, S.; Sagi, I. *J. Am. Chem. Soc.* **2007**, *129*, 13566–13574.

real function of these noncollagenase proteins is still unknown. MMP-2 is able to degrade gelatin, a product of partial hydrolysis of collagen, and MMP-12 elastin, but we do not know if these are their real physiological roles. For MMP-12 it has been reported that the catalytic domain alone is able to degrade elastin even without the hemopexin domain. However, if during evolution the hemopexin domain has been maintained in spite of selective pressure, it is difficult to believe that it is useless. In this respect, the present discovery of relative mobility of the two domains in MMP-12 might be important for this still unknown function.

**Acknowledgment.** We are grateful to Prof. Irit Sagi and to Prof. Christopher M. Overall for the fruitful discussions. This work was supported by EC (Projects: MEST-CT-2004-504391, LSHG-CT-2004-512077, SPINE II LSHG-CT-2006-031220, ORTHO AND PARA WATER no 005032 and Nano4Drugs no LSHB-CT-2005-

019102), by MIUR (PRIN 2005, Prot. N. 2005039878, Prot. RBLA032ZM7), and Ente Cassa di Risparmio di Firenze. We acknowledge the support and assistance of the DESY (Hamburg) synchrotron radiation facilities for the SAXS data collection.

**Supporting Information Available:** NMR chemical shift values of FL-MMP-12 and of the hemopexin domain of MMP-12. The list of upper experimental constraints used for structure calculations. RDC values of FL-MMP-12. Statistical analysis of the NMR structures of the hemopexin domain of MMP-12. Relaxation data. Data collection, processing statistics, and refinement statistics for the crystallographic data set; Figure S1 and S2. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA710491Y

Complementary results :

*Pseudo Contact Shifts observed in presence of Cobalt on MMP12*

Catalytic domain contains a zinc metal ion in its active site. This zinc ion can be substituted by a paramagnetic ion such as cobalt without changing the protein structure for structural studies. From the catalytic domain containing the paramagnetic ion, 59 significant PCS were found (table 1). Considering the hypothesis of a rigid FL-MMP1 protein, simulated PCS on HPX domain from FL-MMP1 X-ray structure were calculated. However, none of these PCS were observed experimentally. As a result, HPX domain is probably moving respect with the CAT domain which contains the metal ion.

num	res	HN	N	num	res	HN	N
110	ARG	0.14	0.17	178	GLY	-0.11	-0.04
113	TYR	0.20	0.31	179	GLY	0.17	0.21
114	ILE	0.24	0.27	188	GLY	-0.03	-0.22
115	THR	0.17	0.15	189	SER	-0.20	-0.17
116	TYR	0.16	0.20	190	GLY	-0.09	-0.19
119	ASN	-0.09	-0.03	191	ILE	-0.25	-0.37
120	ASN	-0.11	-0.26	192	GLY	-0.22	-0.19
121	TYR	-0.14	-0.24	201	GLU	-0.38	-0.57
126	ASN	-0.09	-0.07	203	TRP	-0.42	-0.47
127	ARG	-0.07	-0.01	204	THR	-0.36	-0.35
129	ASP	-0.05	-0.02	205	THR	-0.15	-0.11
130	VAL	-0.04	-0.16	206	HIS	-0.22	-0.14
133	ALA	0.05	0.06	209	GLY	-0.26	-0.18
134	ILE	0.11	0.16	210	THR	-0.16	-0.32
138	PHE	0.58	0.72	212	LEU	-0.30	-0.36
141	TRP	0.77	1.03	213	PHE	-0.18	-0.28
142	SER	0.62	0.58	214	LEU	-0.22	-0.24
143	ASN	0.45	0.49	227	GLY	0.45	0.50
144	VAL	0.47	0.43	231	ASP	-0.87	-0.64
145	THR	0.38	0.26	234	ALA	-0.15	-0.38
150	SER	0.22	0.13	241	LYS	-1.26	-1.23
151	LYS	0.14	0.26	243	VAL	-0.28	-0.16
157	ALA	-0.03	-0.03	248	PHE	0.08	0.05
158	ASP	0.07	0.02	249	ARG	0.13	0.18
162	VAL	-0.34	-0.37	251	SER	0.36	0.29
163	PHE	-0.20	-0.21	257	GLY	0.24	0.13
164	ALA	-0.32	-0.73	260	SER	0.20	0.18
165	ARG	-0.22	-0.27	261	LEU	0.26	0.22
169	GLY	-0.35	-0.31	263	GLY	0.21	0.14
176	GLY	-0.05	-0.04				

**Table 1.** Pseudo contact shift observed from the CAT domain of FL-MMP12 (in ppm)

# Interdomain Flexibility in Full-length Matrix Metalloproteinase-1 (MMP-1)\*<sup>§</sup>

Received for publication, December 22, 2008, and in revised form, February 26, 2009 Published, JBC Papers in Press, March 12, 2009, DOI 10.1074/jbc.M809627200

Ivano Bertini<sup>‡§1</sup>, Marco Fragai<sup>‡¶</sup>, Claudio Luchinat<sup>‡¶</sup>, Maxime Melikian<sup>‡</sup>, Efstratios Mylonas<sup>||</sup>, Niko Sarti<sup>‡§</sup>, and Dmitri I. Svergun<sup>||\*\*</sup>

From the <sup>‡</sup>Magnetic Resonance Center, University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy, the <sup>§</sup>Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy, the <sup>¶</sup>Department of Agricultural Biotechnology, University of Florence, Via Maragliano, 75-77, 50144 Florence, Italy, the <sup>||</sup>European Molecular Biology Laboratory, Hamburg Outstation, Notkestrasse 85, 22603 Hamburg, Germany, and the <sup>\*\*</sup>Institute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, 117333 Moscow, Russia

The presence of extensive reciprocal conformational freedom between the catalytic and the hemopexin-like domains of full-length matrix metalloproteinase-1 (MMP-1) is demonstrated by NMR and small angle x-ray scattering experiments. This finding is discussed in relation to the essentiality of the hemopexin-like domain for the collagenolytic activity of MMP-1. The conformational freedom experienced by the present system, having the shortest linker between the two domains, when compared with similar findings on MMP-12 and MMP-9 having longer and the longest linker within the family, respectively, suggests this type of conformational freedom to be a general property of all MMPs.

Matrix metalloproteinases (MMP)<sup>2</sup> are extracellular hydrolytic enzymes involved in a variety of processes including connective tissue cleavage and remodeling (1–3). All 23 members of the family are able to cleave simple peptides derived from connective tissue components such as collagen, gelatin, elastin, etc. A subset of MMPs is able to hydrolyze more resistant polymeric substrates, such as cross-linked elastin, and partially degraded collagen forms, such as gelatin and type IV collagens (4). Intact triple helical type I–III collagen is only attacked by collagenases MMP-1, MMP-8, and MMP-13 and by MMP-2 and MMP-14 (5–12). Although the detailed mechanism of cleavage of single chain peptides by MMP has been largely elucidated (13–19), little is known about the process of hydrolysis of triple helical collagen. In fact, triple helical collagen cannot

be accommodated in the substrate-binding groove of the catalytic site of MMPs (9).

All MMPs (but MMP-7) in their active form are constituted by a catalytic domain (CAT) and a hemopexin-like domain (HPX) (20–22). The CAT domain contains two zinc ions and one to three calcium ions. One zinc ion is at the catalytic site and is responsible for the activity, whereas the other metal ions have structural roles. The isolated CAT domains retain full catalytic activity toward simple peptides and single chain polymeric substrates such as elastin, whereas hydrolysis of triple helical collagen also requires the presence of the HPX domain (9, 23–25). It has been shown that the isolated CAT domain regains a small fraction of the activity of the full-length (FL) protein when high amounts of either inactivated full-length proteins or isolated HPX domains are added to the assay solution (9). Finally, it has been shown that the presence of the HPX domain alone alters the CD spectrum of triple helical collagen in a way that suggests its partial unwinding (26, 27). It is tempting to speculate that full-length collagenases attack collagen by first locally unwinding the triple helical structure with the help of the HPX domain and then cleaving the resulting, exposed, single filaments (9, 28).

Until 2007, three-dimensional structures of full-length MMPs had been reported only for collagenase MMP-1 (29–31) and gelatinase MMP-2 (32). The structures of the two proteins are very similar and show a compact arrangement of the two domains, which are connected by a short linker (14 and 20 amino acids, respectively). It is difficult to envisage that rigid and compact molecules of this type can interact with triple helical collagen in a way that can lead to first unwinding and then cleavage of individual filaments. It has been recently suggested that such concerted action could occur much more easily if the two domains could enjoy at least a partial conformational independence (9). Slight differences in the reciprocal orientation of the CAT and HPX domains of MMP-1 in the presence (29) and absence (30, 31) of the prodomain were indeed taken as a hint that the two domains could experience relative mobility (29).

Two recent solution studies have shown that conformational independence is indeed occurring in gelatinase MMP-9 (33) and elastase MMP-12 (34), whereas the x-ray structure of the latter (34) is only slightly less compact than those of MMP-1 (29–31) and MMP-2 (32). Among MMPs, MMP-9 features an

\* This work was supported by grants from the European Commission (Projects MEST-CT-2004-504391, SFMET Number 201640, SPINE2-COMPLEXES Number 031220, and Nano4Drugs Number LSHB-CT-2005-019102), the Ministero dell'Istruzione, dell'Università e della Ricerca (PRIN 2005, Protocollo Number 2005039878, Protocollo RBLA032ZM7, Protocollo RBIP06LSS2), and the Ente Cassa di Risparmio di Firenze.

<sup>§</sup> The on-line version of this article (available at <http://www.jbc.org>) contains five supplemental tables.

<sup>1</sup> To whom correspondence should be addressed: via L. Sacconi 6, 50019 Sesto Fiorentino, Italy, Tel.: 39-0554574272; Fax: 39-0554574271. E-mail: [ivanobertini@cerm.unifi.it](mailto:ivanobertini@cerm.unifi.it).

<sup>2</sup> The abbreviations used are: MMP, matrix metalloproteinase; SAXS, small angle x-ray scattering; CAT, catalytic domain; HPX, hemopexin-like domain; FL, full-length; NNGH, *N*-isobutyl-*N*-[4-methoxyphenylsulfonyl]glycyl hydroxamic acid; TROSY, transverse relaxation optimized spectroscopy; HSQC, heteronuclear single quantum correlation; NOE, nuclear Overhauser effect; EOM, ensemble optimization method; DTPA, diethylene-triaminepentaacetic acid; BMA, bismethylamide.

## Interdomain Flexibility in Full-length MMP-1

exceptionally long linker (68 amino acid) (33, 35), which in fact constitutes a small domain by itself (the O-glycosylated domain) (33), and therefore, this inspiring observation can hardly be taken as evidence that conformational freedom is a general characteristic of the two-domain MMPs. MMP-12 features a much more normal 16-amino acid linker, thereby making more probable a general functional role for this conformational freedom (34). However, both MMP-9 and MMP-12 retain their full catalytic activity against their substrates even when deprived of the HPX domain (9). Therefore, the question remains of whether conformational freedom is also a required characteristic for those MMPs that are only active as full-length proteins, *i.e.* collagenases. Interestingly, the three collagenases (MMP-1, MMP-8, and MMP-13) have the shortest linker (14 amino acids) among all MMPs. Demonstrating or negating the presence of conformational freedom in one of these collagenases would therefore constitute a significant step forward to formulate mechanistic hypotheses on their collagenolytic activity.

Our recent studies on MMP-12 in solution (34) have shown that a combination of NMR relaxation studies and small angle x-ray scattering (SAXS) is enough to show the presence and the extent of the relative conformational freedom of the two domains of MMPs. Here we apply the same strategy to full-length MMP-1 and show that sizable conformational freedom is indeed experienced even by this prototypical collagenase, although somewhat less pronounced than that observed for MMP-12.

### EXPERIMENTAL PROCEDURES

**Preparation of Protein Samples**—The cDNA encoding the sequence (Asn-106–Asn-469) of the MMP-1 full-length protein was amplified from TrueClone cDNA (OriGene) by PCR and cloned into the pET21 (Novagen) expression vector using NdeI and XhoI (New England Biolabs) as restriction enzymes. One additional methionine at position 105 was present in the final expression product. The recombinant vector was transformed into *Escherichia coli* strain BL21(DE3) CodonPlus RIPL (Stratagene), and colonies were selected for ampicillin and chloramphenicol resistance. The bacteria were grown in 2×YT medium. When a cell density corresponding to 0.6 A was reached, the expression of the protein was induced by adding 0.5 mM isopropyl-β-D-thiogalactoside, and the incubation at 37 °C was continued for another 5 h. The full-length MMP-1 precipitated in the inclusion bodies, and these were solubilized, after lysis of the cells, in a solution of 8 M urea, 20 mM dithiothreitol, and 20 mM Tris-HCl, (pH 8.2). The protein was diluted with a buffer containing 6 M urea, 10 mM CaCl<sub>2</sub>, 0.1 mM ZnCl<sub>2</sub>, 20 mM cystamine, 20 mM Tris-HCl (pH 8) and refolded by using a multistep dialysis against solutions containing 50 mM Tris-HCl (pH 8), 4 M urea, 10 mM CaCl<sub>2</sub>, 0.1 mM ZnCl<sub>2</sub>, 5 mM β-mercaptoethanol, 1 mM 2-hydroxyethyl disulfide, then against a solution containing 50 mM Tris-HCl (pH 7.2), 2 M urea, 10 mM CaCl<sub>2</sub>, 0.1 mM ZnCl<sub>2</sub>, 0.3 M NaCl, and then against the same solution without urea. The protein was purified by size-exclusion chromatography on the HiLoad 26/60 Superdex 75 pg (Amersham Biosciences). For expression of <sup>15</sup>N- and <sup>13</sup>C-enriched FL-MMP-1, the bacteria were grown in minimal

medium containing <sup>15</sup>N-enriched (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and <sup>13</sup>C-enriched glucose (Cambridge Isotope Laboratories). The resulting inclusion bodies were solubilized in 8 M urea, 20 mM Tris-HCl (pH 8), 20 mM dithiothreitol. Protein molecular weight and purity were checked on a 17% gel by SDS-PAGE and by mass spectroscopy (matrix-assisted laser desorption). The E219A mutant of FL-MMP-1 was produced using the QuikChange site-directed mutagenesis kit (Qiagen), and the expression and purification of the protein and of its <sup>15</sup>N- and <sup>13</sup>C-<sup>15</sup>N-enriched versions were completed using the same procedure described above. Samples of <sup>2</sup>H-<sup>13</sup>C-<sup>15</sup>N-enriched full-length human MMP-1 protein were obtained by adapting *E. coli* cells in medium with different percentages of deuterium until reaching 100%, and then growing the cells in OD2 Silantes media carbon, deuterium, nitrogen (CDN). Samples of cadmium(II)-substituted FL-MMP-1 protein were prepared by exhaustive dialysis against a buffer containing 20 mM Tris, pH 7.2, 10 mM CaCl<sub>2</sub>, 0.3 M NaCl, 0.2 M acetohydroxamic acid, and 0.3 mM CdCl<sub>2</sub> (36). Equimolar concentrations of *N*-isobutyl-*N*-[4-methoxyphenyl-sulfonyl] glycyl hydroxamic acid (NNGH) were added to the samples to further increase the protein stability. The cDNA encoding for the HPX domain (Thr-274–Asn-469) was generated by polymerase chain reaction and cloned into pET21a using NdeI and XhoI as restriction enzymes. The expression vector was then transformed into competent *E. coli* BL21(DE3) CodonPlus RIPL strain, and the colonies were selected for ampicillin and chloramphenicol resistance. Protein refolding both for non-enriched samples and for <sup>13</sup>C- and/or <sup>15</sup>N-enriched samples was carried out by using the same protocols described previously for the preparation of FL-MMP-1 samples. Samples of the zinc(II) catalytic domain (Asn-106–Gly-261, E219A inactive mutant) were prepared as described previously (37).

**NMR Measurements and Protein Assignment**—The experiments for the protein assignment and mobility measurements of the isolated domains (CAT and HPX) were performed on protein samples at concentrations ranging between 0.3 and 0.7 mM (pH 7.2). For FL-MMP-1, all NMR experiments were performed on samples at a concentration of 0.3 mM (pH 7.2). NMR experiments were performed at 298 and 306 K and acquired on Bruker AVANCE 900, AVANCE 800, AVANCE 700, AVANCE 600, and DRX 500 spectrometers equipped with triple resonance cryo-probes. All spectra were processed with the Bruker TOPSPIN software packages and analyzed by the program CARA (Computer Aided Resonance Assignment, ETH Zürich) (38)). The backbone resonance assignment of HPX domain was obtained by the analysis of HNCA, HNCACB, and CBCA-(CO)NH spectra performed at 500 and 900 MHz, whereas the assignment of the <sup>2</sup>H-, <sup>15</sup>N-, and <sup>13</sup>C-enriched FL-MMP-1 was obtained by the analysis TROSY-HNCA, TROSY-HNCACB TROSY-HNCOACB performed on an 800-MHz spectrometer. The obtained assignments are reported in supplemental Table S1 for the full-length protein, in supplemental Table S2 for the catalytic domain, and in supplemental Table S3 for the hemopexin-like domain.

The protein assignment and the mobility measurements on FL-MMP-1 were performed on the NNGH-inhibited, E219A mutant, due to its high stability to the self-hydrolysis. Mobility

measurements on the catalytic domain were performed on the NNGH-inhibited form of the protein.

**$R_1$ ,  $R_2$ , and NOE Measurements**—The experiments for the determination of  $^{15}\text{N}$  longitudinal and transverse relaxation rates and  $^1\text{H}$ - $^{15}\text{N}$  NOE were recorded at 306 K and 700 MHz on  $^{15}\text{N}$ -enriched samples. The  $^{15}\text{N}$  longitudinal relaxation rates ( $R_1$ ) were measured using a sequence modified to remove cross correlation effects during the relaxation delay (39). Inversion recovery times ranging between 2.5 and 3000 ms, with a recycle delay of 3.5 s, were used for the experiments. The  $^{15}\text{N}$  transverse relaxation rates ( $R_2$ ) were measured using a Carr-Purcell-Meiboom-Gill (CPMG) sequence (39, 40) with delays ranging between 8.5 and 237.4 ms for the CAT domain, between 8.5 and 203.5 ms for the HPX domain, and finally between 8.5 and 135.7 ms for the FL-MMP-1 protein with a refocusing delay of 450  $\mu\text{s}$ . The relaxation data are reported in supplemental Table S4.  $R_1$  and  $R_2$  data measured on the full-length protein were found noisier and less uniform with respect to those of the single catalytic and hemopexin-like domains. This is related to the overlap of the signals in such a large protein and to the relative low solubility of the full-length construct.

**Paramagnetic Relaxation Enhancement Measurements**—The paramagnetic relaxation enhancements of the backbone NH protons were obtained by adding a stock water solution (50 mM) of Gd(DTPA·BMA) to a solution of FL-MMP-1 (200  $\mu\text{M}$  protein in a buffer containing 20 mM Tris, pH 7.2, 10 mM  $\text{CaCl}_2$ , 0.3 M NaCl, 0.2 M acetohydroxamic acid, 1 mM  $\text{ZnCl}_2$ , and 3 mM NNGH) up a final concentration of 1.4 mM. NH proton longitudinal relaxation times were measured through an inversion recovery HSQC sequence obtained by introducing a  $^1\text{H}$  180° pulse followed by a variable delay  $\tau$  in front of a standard  $^1\text{H}$ - $^{15}\text{N}$ -HSQC sequence.  $T_1$  values were obtained from a series of spectra obtained with the following  $\tau$  values (in ms): 20, 50, 80, 100, 200, 400, 500, 800, 1000, 1400, 2000, 2200, 3000, and 3500. The measurement was repeated on the free protein and after each addition of the paramagnetic complex. Peak volumes as a function of the  $\tau$  value were fitted to a monoexponential recovery with a three-parameter fit.

**SAXS Experiments and Data Analysis**—Small Angle x-ray scattering data from solutions of the NNGH-inhibited, cadmium(II)-substituted E219A mutant of FL-MMP-1 were collected on the X33 beamline of the European Molecular Biology Laboratory (EMBL) (Deutsches Elektronen Synchrotron (DESY), Hamburg) (41) using a MAR345 image plate detector. The scattering patterns were measured with a 2-min exposure time for several solute concentrations in the range from 0.8 to 8.3 mg/ml. To check for radiation damage, two 2-min exposures were compared, and no changes were detected. Using the sample-detector distance of 2.7 m, a range of momentum transfer of  $0.01 < s < 0.5 \text{ \AA}^{-1}$  was covered ( $s = 4\pi \sin(\theta)/\lambda$ , where  $2\theta$  is the scattering angle and  $\lambda = 1.5 \text{ \AA}$  is the x-ray wavelength). The data were processed using standard procedures and extrapolated to infinite dilution using the program PRIMUS (42). The forward scattering,  $I(0)$ , and the radius of gyration,  $R_g$ , were evaluated using the Guinier approximation (43), assuming that at very small angles ( $s < 1.3/R_g$ ), the intensity is represented as  $I(s) = I(0)\exp((-s^2R_g^2)/3)$ . The values of  $I(0)$  and  $R_g$ , as well as the maximum dimension,  $D_{\text{max}}$ , and the interatomic distance

distribution functions,  $(p(r))$ , were also computed using the program GNOM (44). The scattering from the high resolution models was computed using the program CRY SOL (45). Given the atomic coordinates, the program predicts the theoretical scattering pattern and fits the experimental intensity by adjusting the excluded volume of the particle and the contrast of the hydration layer to minimize the discrepancy

$$\chi^2 = \frac{1}{N-1} \sum_j \left[ \frac{I_{\text{exp}}(s_j) - cI_{\text{calc}}(s_j)}{\sigma(s_j)} \right]^2$$

where  $N$  is the number of experimental points,  $c$  is a scaling factor,  $I_{\text{exp}}(s_j)$  and  $I_{\text{calc}}(s_j)$  are the experimental and calculated intensities, respectively, and  $\sigma(s_j)$  is the experimental error at the momentum transfer  $s_j$ . To assess the conformational variability of MMP-1, an ensemble optimization method (EOM) was used (46), allowing for the coexistence of multiple conformations in solution. About 10,000 randomized models of FL-MMP-1 differing by the conformation of the interdomain linker were generated, and their scattering patterns were computed using the program RanCh of the EOM package. These models formed a pool of possible structures for which the scattering patterns were computed by CRY SOL. The EOM program employs a genetic algorithm to select from the pool a small number (usually about 20) of representative structures such that the average scattering from the selected ensemble fits the experimental data. Multiple runs of EOM were performed, and the results were averaged to provide quantitative information about the flexibility of the protein in solution (in particular, about the  $R_g$  distribution in the selected ensembles).

## RESULTS AND DISCUSSION

Full-length human MMP-1 and its CAT-(106–261) and HPX-(274–469) domains were expressed as described under “Experimental Procedures.” In both FL-MMP-1 and its CAT domain, an E219A mutation was introduced. Such mutation ensures a dramatic lowering of the catalytic activity (13, 19), which is needed in the present investigation to prevent self-hydrolysis. In addition, the strong active site-directed hydroxamic inhibitor NNGH (15) was always added. Finally, when needed, cadmium(II) was substituted for zinc(II) in the catalytic site to further reduce the residual activity as it was already shown for FL-MMP-12 and its CAT domain (34).  $^{15}\text{N}$ -,  $^{13}\text{C}$ - $^{15}\text{N}$ -, and  $^2\text{H}$ - $^{13}\text{C}$ - $^{15}\text{N}$ -enriched samples were used for NMR investigations. The samples were stable under the NMR experimental conditions for several months. For the present mobility studies, the assignment of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of the three constructs was needed. The two-dimensional and three-dimensional experiments that were used for this purpose are described under “Experimental Procedures.” The  $^1\text{H}$ - $^{15}\text{N}$ -TROSY-HSQC spectra of the three constructs, acquired at 306 K, are superimposed in Fig. 1. The spectral quality, especially for the FL and HPX domains, is better at 306 K than at 298 K. This prompted us to perform all subsequent work at 306 K. Moreover, the latter temperature is closer to the physiological value of 310 K, so the resulting mobility data will be more meaningful.



## Interdomain Flexibility in Full-length MMP-1

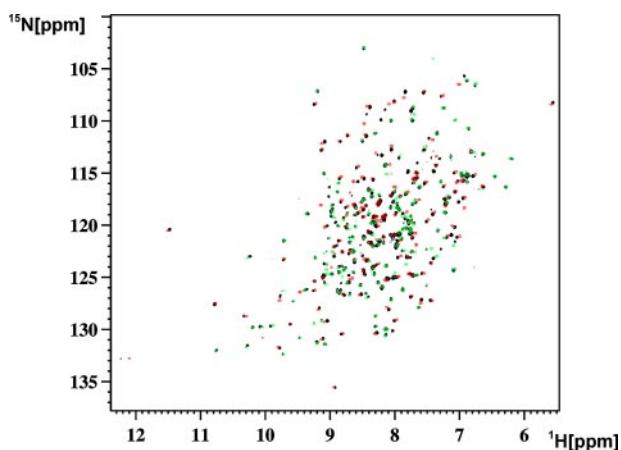


FIGURE 1.  $^1\text{H}$ - $^{15}\text{N}$ -TROSY-HSQC spectra of FL-MMP-1 ( $^2\text{H}$ - $^{13}\text{C}$ - $^{15}\text{N}$ -labeled NNGH-inhibited, E219A mutant) (black), superimposed with CAT-MMP-1 ( $^{15}\text{N}$ -labeled NNGH-inhibited, E219A mutant) (red) and with HPX-MMP-1 ( $^{15}\text{N}$ -labeled) (green).

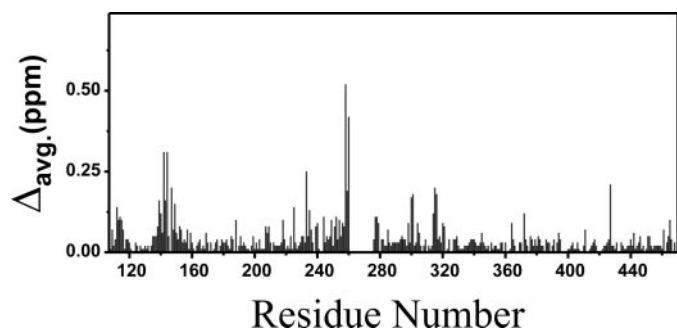


FIGURE 2. Plot of mean shift differences experienced by FL MMP-1 with respect to the isolated CAT or HPX domains. *avg.* indicates average.

*Analysis of the Chemical Shift Differences between FL-MMP-1 and Its Isolated Domains*—For clarity, the spectrum of the full-length protein in Fig. 1 is from a triply labeled  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  sample, which displays much sharper lines. From visual inspection, it already appears that the FL spectrum is largely the superposition of the CAT and HPX spectra as the majority of the signals in the FL spectrum overlap with either a CAT or an HPX signal. There are a number of additional signals in the FL spectrum that are attributed to the portion of the interdomain linker (262–275), which is missing in both the CAT and the HPX isolated domains. Furthermore, a few more signals undergo a modest shift on passing from the isolated domains to the FL protein. These signals could arise from contacts between each domain and the linker, contacts between the two domains, or both. In the latter two cases, evidence of interdomain contacts would imply that either the FL length structure is rigid or, that even in the presence of conformational freedom, “closed” compact structures contribute to the overall description of the molecule.

The assignments for the three constructs are reported in supplemental Tables S1–S3. From a comparison of the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts in the  $^1\text{H}$ - $^{15}\text{N}$ -TROSY-HSQC spectra for the FL and the two isolated domains, a mean shift difference plot (47) was generated (Fig. 2). It appears that the number of peaks experiencing chemical shift differences is rather limited and clustered in the 112–115, 139–149, 244–261, and 277–316 regions. Fig. 3A shows the experimental x-ray three-dimen-

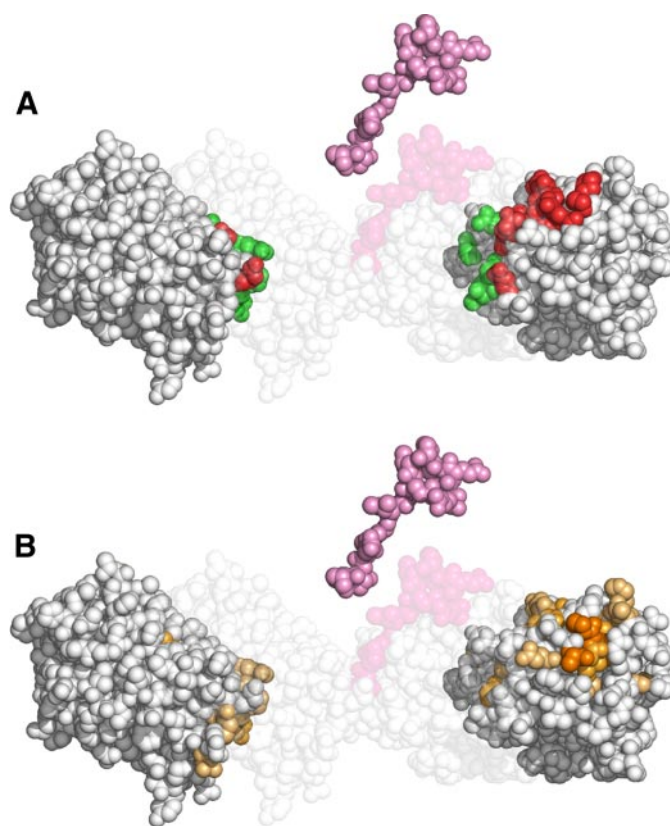
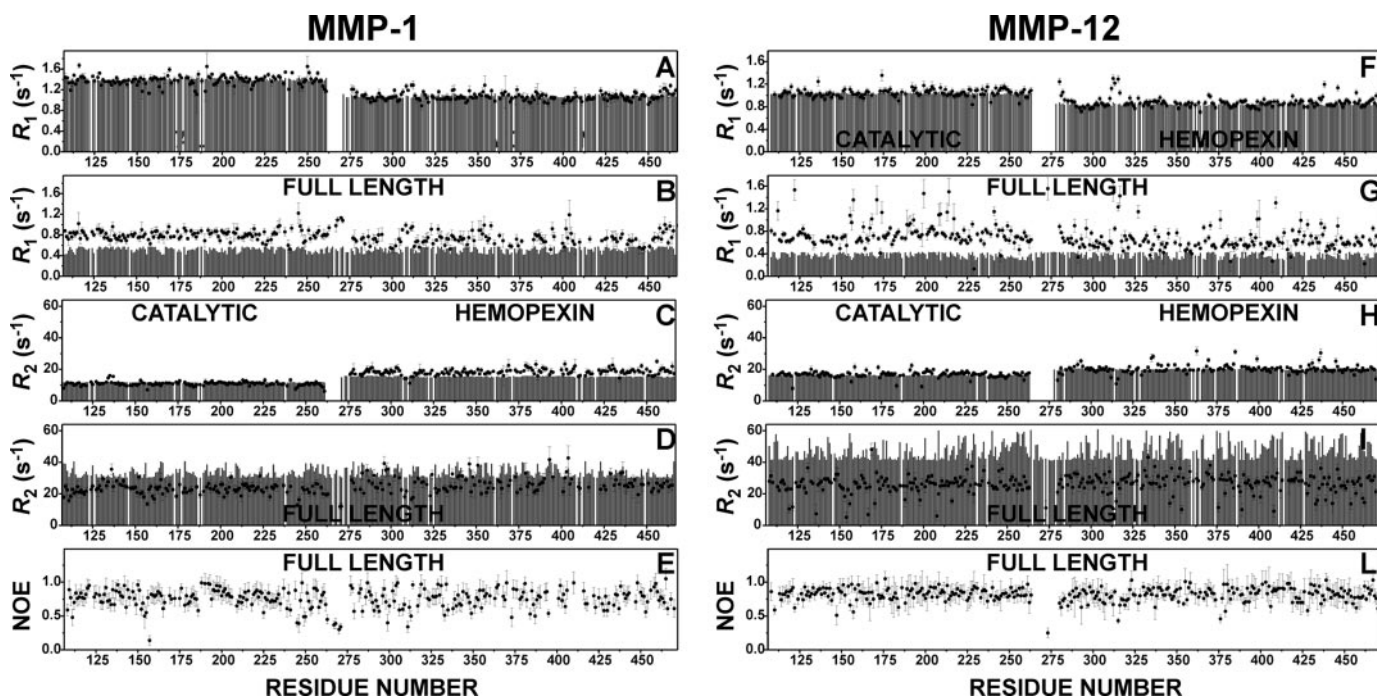


FIGURE 3. Mapping of the residues at the interface between the CAT domain, the linker and the HPX domain. *A*, exploded space fill representation of the experimental three-dimensional x-ray structure of FL-MMP-1 (24) where the regions of contact between CAT and HPX domains are colored in green and the regions of contact of either of the two domains with the linker in the crystal structure is visible in the background. *B*, the same exploded space fill representation where the intensity of the colored residues, from yellow to dark-orange, correlates with the size of the mean shift difference experienced by FL-MMP-1 with respect to the isolated domains. The similarity of the areas experiencing shift changes due to the lack of the linker in *B* with the contact areas in red, but not with those in green in *A*, show that most of the shift perturbation arises from the contacts with the linker and not between the two domains.

sional structure of the FL protein (30) as well as of its isolated domains and linker, exploded to show the regions of contact of the CAT and HPX domains with one another (green) and with the linker (red). Fig. 3B shows the same structures, where the intensity of the colored residues correlates with the size of the mean shift difference experienced by the FL with respect to the isolated domains. It is apparent that most of the chemical shift differences arise from contacts of the CAT and HPX domains with the linker, and very few and with relatively low values are attributable to direct interdomain contacts. It can be concluded that it is likely that the compact structure observed in the x-ray structure of the active FL protein is in equilibrium with other more open structures in solution.

*Relaxation Measurements*—Measurements of longitudinal ( $R_1$ ) and transverse ( $R_2$ ) relaxation rates of backbone amide nitrogens at 700-MHz  $^1\text{H}$  Larmor frequency and 306 K have been performed on  $^{15}\text{N}$ -enriched samples of both the isolated CAT domains and the isolated HPX domains as well as of FL-MMP-1. Estimates of  $R_1$  and  $R_2$  values for these three constructs under the chosen experimental conditions of magnetic



**FIGURE 4. Comparison of NMR relaxation data for MMP-1 (data collected at 306 K) and MMP-12 (data collected at 298 K).** The calculated (gray bars) and experimental (filled circles) backbone  $^{15}\text{N}$   $R_1$  and  $R_2$  values for the isolated CAT and HPX domains of MMP-1 are shown in A and C, and those of MMP-12 are shown in F and H. The  $^{15}\text{N}$   $R_1$  and  $R_2$  values for the full-length MMP-1 and MMP-12 are shown in B, D, G, and I, respectively. The experimental NOE values for the full-length proteins are shown in E and L for MMP-1 and MMP-12, respectively. In both proteins, the agreement between experimental and calculated  $R_1$  values for the isolated domains is excellent, whereas for the full-length proteins, the experimental  $R_1$  values are sizably larger (B and G) than the ones calculated for the rigid x-ray structures.  $R_2$  values (D and I) are also in good agreement if a slight tendency of the HPX domain to aggregate is taken into account. The data comparisons are only consistent with interdomain flexibility in MMP-1, which is somewhat less extensive than in MMP-12.

field and temperature were obtained using the program HydroNMR (48), and the three-dimensional structures of the constructs were taken from the x-ray structure of the FL protein (30). The experimental and theoretical  $R_1$  and  $R_2$  values are reported in Fig. 4, A–E, together with the corresponding data for MMP-12 (Fig. 4, F–L) (34) at the same field and 298 K, for comparison purposes. If the experimental data do not agree with the calculated data, outside the experimental error, then they are inconsistent with the structural model. As far as the catalytic domain is concerned (left side), it is clear that the  $R_1$  and  $R_2$  data for the isolated domain are in excellent agreement with the theoretical expectations. On the contrary, the experimental  $R_1$  data for the CAT domain, when it is part of the FL protein, are sizably higher, and the  $R_2$  data are sizably lower, with respect to the theoretical values for a rigid FL protein. Higher  $R_1$  and lower  $R_2$  values, taken together, indicate that the two domains behave as if they belonged to a lower molecular weight protein; that is, the CAT domain possesses some degree of motion, which is independent of the motion of the FL protein as a whole.

In principle, the same should hold for the HPX domain. Indeed, the  $R_1$  and  $R_2$  data for the HPX domain in the full-length protein are also higher and lower, respectively, than the theoretical estimate, but here the difference is less marked than for the CAT domain. A plausible explanation lies in the fact that the  $R_1$  and  $R_2$  data for the isolated HPX domain show the opposite behavior, *i.e.* the  $R_1$  values are lower and the  $R_2$  values are higher than expected. This is typically encountered in the presence of self-aggregation. The HPX domain does possess a

rather hydrophobic surface, which could allow for the existence of a fraction of dimeric or higher order aggregation species in solution. Partial aggregation brought about by the HPX domain in the FL protein would of course artificially attenuate the effect of interdomain mobility on  $R_1$  and  $R_2$ . The hypothesis of HPX aggregation is easy to check by changing the protein concentration. Indeed, for isolated HPX solutions of higher concentration, the discrepancy with the calculated data sizably increases (data not shown). In conclusion, the HPX data (i) demonstrate the presence of aggregation but also (ii) confirm interdomain mobility in the full-length protein, which is still apparent from  $R_1$  and  $R_2$  data despite the attenuation of the effect due to aggregation.

A further proof of the interdomain mobility is provided by the very small NOE values of four adjacent residues in the linker region (267, 268, 270, and 271) (Fig. 4E). Comparison with the MMP-12 data (Fig. 4, F–L), collected at lower temperature, shows that interdomain mobility in MMP-1, although certainly present, is less extensive than in MMP-12. This is possibly related (i) to the shorter linker (14 *versus* 16 residues) and (ii) to the more extensive interdomain contact present in the x-ray structure of active FL-MMP-1 with respect to that observed in the x-ray structure of MMP-12.

*A Paramagnetic Probe Monitors Interdomain Surface Exposure*—Paramagnetic probes dissolved in protein solutions have been previously used (i) to identify surface residues to help structure determination (49–56), (ii) to determine the orientation of membrane-bound peptides on the membrane surface (57–59), and (iii) to monitor the formation of protein-protein

## Interdomain Flexibility in Full-length MMP-1

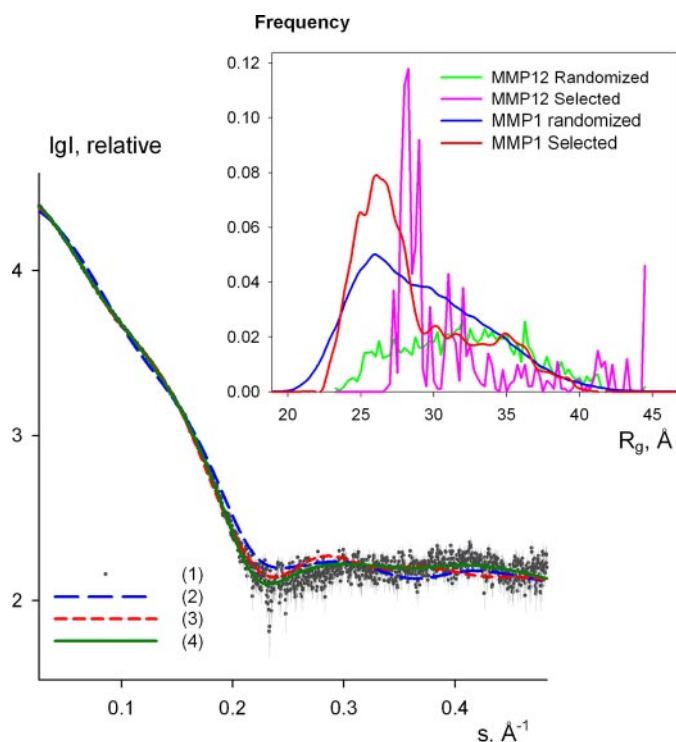


FIGURE 5. Experimental x-ray scattering from the NNGH-inhibited, FL-MMP-1 and scattering calculated from the following models: *curve 1*, experimental data with error bars; *curves 2 and 3*, computed scattering from the crystallographic models of FL-MMP-1 and FL-MMP-12, respectively; and *curve 4*, a typical fit by the selected ensemble of structures. The logarithm of the scattering intensity is plotted against the momentum transfer. *Inset*, the frequency of the models with the given  $R_g$  in the initial pool of structures with randomized interdomain linkers (FL-MMP-1 (blue) and FL-MMP-12 (magenta)) and in the selected ensembles (FL-MMP-1 (red) and FL-MMP-12 (green)); the latter distribution is obtained by the averaging of several EOM runs. All histograms are normalized to the integral value of unity.

complexes by detecting residues that become shielded from the probe upon complex formation (4, 60, 61). In analogy with the latter strategy, we have used the gadolinium complex Gd(DTPA-BMA) (57) to probe the CAT-HPX interface in the FL protein. An inversion recovery HSQC sequence was used to measure the  $R_1$  values of the backbone NH protons of FL-MMP-1 in the absence ( $R_{1\text{dia}}$ ) and in the presence ( $R_{1\text{para}}$ ) of 1.4 mM Gd(DTPA-BMA), as detailed under “Experimental Procedures.” The residues for which the relaxation rate difference  $R_{1\text{para}} - R_{1\text{dia}}$  is higher than a threshold value of  $0.7 \text{ s}^{-1}$  are reported in supplemental Table S5. Among them, on the HPX side, residue Leu-314 should be shielded by the CAT domain; on the CAT side, residue Asp-231 should be shielded by the linker, and residues Asp-245, Val-246, and Gln-247 should be shielded by the HPX domain. On the contrary, all of them experience paramagnetic effects particularly strong for Asp-245 ( $2.7 \text{ s}^{-1}$ ) and Val-246 ( $2.4 \text{ s}^{-1}$ ). These effects are inconsistent with the compact x-ray structure of FL-MMP-1 and confirm that the protein must spend a non-negligible fraction of time in an extended conformation so that the CAT-HPX interface becomes accessible to the probe.

**SAXS Measurements**—The processed x-ray scattering pattern from FL-MMP-1 is shown in Fig. 5. The experimental radius of gyration  $R_g$  is  $29 \pm 1 \text{ \AA}$ . This value significantly exceeds that calculated ( $R_g = 25 \text{ \AA}$ ) from the x-ray structure of

active FL-MMP-1 (30). Moreover, the scattering pattern calculated from the active FL-MMP-1 model using CRY SOL (45) fails to fit the experimental data (discrepancy  $\chi = 4.22$ , Fig. 5, *curve 2*). Interestingly, the fit to the FL-MMP-12 model is better (discrepancy  $\chi = 1.8$ , Fig. 5, *curve 3*). If a distribution of FL-MMP-1 conformations, including compact and extended ones, is taken into consideration, neither individual models nor averaging over the random pool allowed one to fit the SAXS data satisfactorily ( $\chi > 1.9$  in both cases). The representative ensembles selected to fit the data give information about the preferable conformations of the protein. To assess the preferable conformations in solution, the EOM method (46) was used. Given a representative pool of (random) structures, the method employs a genetic algorithm to select the ensembles from this pool that best fit the experimental data, as explained under “Experimental Procedures.” Several EOM runs yielded reproducible ensembles neatly fitting the experimental data with discrepancy  $\chi$  less than 1.2, and a typical fit provided by the ensemble selected by EOM is given in Fig. 5, *curve 4*. All the fits from different EOM runs are graphically indistinguishable from Fig. 5, *curve 4*. The  $R_g$  distributions of the particles in the initial pool and in the selected ensembles are compared in Fig. 5 (*inset*). The former distribution is rather broad and covers the  $R_g$  range from about 20 to 45 Å, corresponding to extremely compact and completely extended domain configurations, respectively. In contrast, the  $R_g$  distribution of the selected ensembles displays a relatively sharp peak around  $R_g = 25\text{--}26 \text{ \AA}$ , including about 70% of the particles in the selected ensembles. Visual inspection of the models having  $R_g$  values in the peak range indicates, not unexpectedly, that their shape is similar to that of the FL-MMP-1 structure in the crystal (30) (although with varying interdomain orientations). However, an extended tail at higher  $R_g$  values up to around 40 Å, accounting for about 30% of the particles, is apparent from the multiple EOM reconstructions. These results indicate that the crystallographic conformation of FL-MMP-1 (30) is largely present also in solution, but in addition, the protein experiences noticeable flexibility with a significant amount of extended conformations in equilibrium with the closed one(s). We also tried to explore the possibility of a two-state exchange situation allowing for only two conformations in the mixture. In such a case, two structures with an  $R_g$  of around 25 and 32 Å are selected. It is interesting that FL-MMP-1 (30) appears much more compact than FL-MMP-12 (34), as shown by the comparison in Fig. 5, *inset*, indicating that the two extra residues in the linker region of FL-MMP-12 increase its flexibility.

**Concluding Remarks and Biological Implications**—The present data demonstrate that FL-MMP-1 shows relative mobility of its catalytic and hemopexin-like domains, as recently observed for FL-MMP-12 (34). As in the latter case, the reorientation of the backbone NH vectors with respect to the magnetic field occurs on a time scale that is faster than the rotational time of the whole molecule (62). The amplitude of the motion is probably smaller for FL-MMP-1 than it is for FL-MMP-12, as judged from the SAXS data that suggest that the molecules spend two-thirds of the time in a conformation that is more or less as compact as the solid state structure but is ample enough that the residues at the interface between

the two domains are significantly exposed to a paramagnetic probe in solution. Besides MMP-12, relative mobility of the CAT and HPX domains has also been recently suggested for MMP-9, where the linker between the two domains is much longer and constitutes a domain by itself (O-glycosylated domain) (63). In the case of MMP-9, it has been argued (33) that the long O-glycosylated domain may be flexible and may mediate protein-substrate interactions (64, 65). However, the present finding of flexibility in FL-MMP-1 has further biological implications because MMP-1 is a collagenase that is able to hydrolyze intact triple helical type I–III collagen but only in the full-length form, *i.e.* the presence of the HPX domain is crucial for the function. Notably, the possibility of reorienting the HPX with respect to the CAT domain during catalysis has often been invoked as a means to unwind the collagen triple helix in preparation for the catalytic cleavage (66).

*Acknowledgments*—We acknowledge the support and assistance of the DESY (Hamburg) synchrotron radiation facilities for the SAXS data collection.

## REFERENCES

1. Woessner, J. F., Jr., and Nagase, H. (1999) *J. Biol. Chem.* **274**, 21491–21494
2. Brinckerhoff, C. E., and Matrisian, L. M. (2002) *Nat. Rev. Mol. Cell Biol.* **3**, 207–214
3. Page-McCaw, A., Ewald, A. J., and Werb, Z. (2007) *Nat. Rev. Mol. Cell Biol.* **8**, 221–233
4. Bhaskaran, R., Palmier, M. O., Lauer-Fields, J. L., Fields, G. B., and Van Doren, S. R. (2008) *J. Biol. Chem.* **283**, 21779–21788
5. Sternlicht, M. D., and Werb, Z. (2001) *Annu. Rev. Cell Dev. Biol.* **17**, 463–516
6. Overall, C. M. (2002) *Mol. Biotechnol.* **22**, 51–86
7. Lauer-Fields, J. L., Juska, D., and Fields, G. B. (2002) *Biopolymers* **66**, 19–32
8. Visse, R., and Nagase, H. (2003) *Circ. Res.* **92**, 827–839
9. Chung, L. D., Dinakarandian, D., Yoshida, N., Lauer-Fields, J. L., Fields, G. B., Visse, R., and Nagase, H. (2004) *EMBO J.* **23**, 3020–3030
10. Gioia, M., Fasciglione, G. F., Marini, S., D'Alessio, S., De Sanctis, G., Diekmann, O., Pieper, M., Politi, V., Tschesche, H., and Coletta, M. (2002) *J. Biol. Chem.* **277**, 23123–23130
11. Gioia, M., Monaco, S., Fasciglione, G. F., Coletti, A., Modesti, A., Marini, S., and Coletta, M. (2007) *J. Mol. Biol.* **368**, 1101–1113
12. Lauer-Fields, J. L., Tuzinski, K. A., Shimokawa, K.-I., Nagase, H., and Fields, G. B. (2000) *J. Biol. Chem.* **275**, 13282–13290
13. Bertini, I., Calderone, V., Fragai, M., Luchinat, C., and Maletta, M. (2006) *Angew. Chem. Int. Ed. Engl.* **45**, 7952–7955
14. Bertini, I., Calderone, V., Fragai, M., Luchinat, C., Mangani, S., and Terni, B. (2003) *Angew. Chem. Int. Ed. Engl.* **42**, 2673–2676
15. Bertini, I., Calderone, V., Cosenza, M., Fragai, M., Lee, Y.-M., Luchinat, C., Mangani, S., Terni, B., and Turano, P. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102**, 5334–5339
16. Rosenblum, G., Meroueh, S., Toth, M., Fisher, J. F., Fridman, R., Mobashery, S., and Sagi, I. (2007) *J. Am. Chem. Soc.* **129**, 13566–13574
17. Gall, A. L., Ruff, M., Kannan, R., Cuniasso, P., Yiotakis, A., Dive, V., Rio, M. C., Basset, P., and Moras, D. (2001) *J. Mol. Biol.* **307**, 577–586
18. Bode, W., Reinemer, P., Huber, R., Kleine, T., Schnierer, S., and Tschesche, H. (1994) *EMBO J.* **13**, 1263–1269
19. Cha, J., and Auld, D. S. (1997) *Biochemistry* **36**, 16019–16024
20. Massova, I., Kotra, L. P., Fridman, R., and Mobashery, S. (1998) *FASEB J.* **12**, 1075–1095
21. Bode, W., and Maskos, K. (2003) *Biol. Chem.* **384**, 863–872
22. Nagase, H., Visse, R., and Murphy, G. (2006) *Cardiovasc. Res.* **69**, 562–573
23. Clark, I. E., and Cawston, T. E. (1989) *Biochem. J.* **263**, 201–206
24. Murphy, G., Allan, J. A., Willenbrock, F., Cockett, M. I., O'Connell, J. P., and Docherty, A. J. (1992) *J. Biol. Chem.* **267**, 9612–9618
25. Knauper, V., Cowell, S., Smith, B., Lopez-Otin, C., O'Shea, M., Morris, H., Zardi, L., and Murphy, G. (1997) *J. Biol. Chem.* **272**, 7608–7616
26. Tam, E. M., Wu, Y. I., Butler, G. S., Stack, M. S., and Overall, C. M. (2002) *J. Biol. Chem.* **277**, 39005–39014
27. Tam, E. M., Moore, T. R., Butler, G. S., and Overall, C. M. (2004) *J. Biol. Chem.* **279**, 43336–43344
28. Minond, D., Lauer-Fields, J. L., Nagase, H., and Fields, G. B. (2004) *Biochemistry* **43**, 11474–11481
29. Jozic, D., Bourenkov, G., Lim, N. H., Visse, R., Nagase, H., Bode, W., and Maskos, K. (2005) *J. Biol. Chem.* **280**, 9578–9585
30. Iyer, S., Visse, R., Nagase, H., and Acharya, K. R. (2006) *J. Mol. Biol.* **362**, 78–88
31. Li, J., Brick, P., Ohare, M. C., Skarzynski, T., Lloyd, L. F., Curry, V. A., Clark, I. M., Bigg, H. F., Hazleman, B. L., Cawston, T. E., and Blow, D. M. (1995) *Structure (Lond.)* **3**, 541–549
32. Morgunova, E., Tuuttila, A., Bergmann, U., Isupov, M., Lindqvist, Y., Schneider, G., and Tryggvason, K. (1999) *Science* **284**, 1667–1670
33. Rosenblum, G., Van den Steen, P. E., Cohen, S. R., Grossmann, J. G., Frenkel, J., Sertchook, R., Slack, N., Strange, R. W., Opdenakker, G., and Sagi, I. (2007) *Structure (Lond.)* **15**, 1227–1236
34. Bertini, I., Calderone, V., Fragai, M., Jaiswal, R., Luchinat, C., Melikian, M., Mylonas, E., and Svergun, D. (2008) *J. Am. Chem. Soc.* **130**, 7011–7021
35. Andreini, C., Banci, L., Bertini, I., Luchinat, C., and Rosato, A. (2004) *J. Proteome Res.* **3**, 21–31
36. Bertini, I., Fragai, M., Lee, Y.-M., Luchinat, C., and Terni, B. (2004) *Angew. Chem. Int. Ed. Engl.* **43**, 2254–2256
37. Nesi, A., and Fragai, M. (2007) *ChemBioChem* **8**, 1367–1369
38. Keller, R. L. J. (2004) *The Computer Aided Resonance Assignment Tutorial*, Vol. 1.3, CANTINA Verlag, Goldau, Switzerland
39. Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A., and Torchia, D. A. (1992) *J. Magn. Reson.* **97**, 359–375
40. Peng, J. W., and Wagner, G. (1994) *Methods Enzymol.* **239**, 563–596
41. Roessle, M. W., Klaering, R., Ristau, U., Robrahn, B., Jahn, D., Gehrman, T., Konarev, P. V., Round, A., Fiedler, S., Hermes, S., and Svergun, D. I. (2007) *J. Appl. Crystallogr.* **40**, s190–s194
42. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) *J. Appl. Crystallogr.* **36**, 1277–1282
43. Guinier, A. (1939) *Ann. Phys.* **12**, 161–237
44. Svergun, D. I. (1992) *J. Appl. Crystallogr.* **25**, 495–503
45. Svergun, D. I., Barberato, C., and Koch, M. H. J. (1995) *J. Appl. Crystallogr.* **28**, 768–773
46. Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007) *J. Am. Chem. Soc.* **129**, 5656–5664
47. Grzesiek, S., Bax, A., Clore, G. M., Gronenborn, A. M., Hu, J. S., Kaufman, J., Palmer, I., Stahl, S. J., and Wingfield, P. T. (1996) *Nat. Struct. Biol.* **3**, 340–345
48. de la Torre, J. G., Huertas, M. L., and Carrasco, B. (2000) *J. Magn. Reson.* **147**, 138–146
49. Aime, S., D'Amelio, N., Fragai, M., Lee, Y.-M., Luchinat, C., Terreno, E., and Valensin, G. (2002) *J. Biol. Inorg. Chem.* **7**, 617–622
50. Pintacuda, G., and Otting, G. (2002) *J. Am. Chem. Soc.* **124**, 372–373
51. Scarselli, M., Esposito, G., De Magistris, M. T., Domenighini, M., Rappuoli, R., Burrioni, G., Bernini, A., and Niccolai, N. (1998) *J. Biochem.* **254**, 313–317
52. Scarselli, M., Bernini, A., Segoni, C., Molinari, H., Esposito, G., Lesk, A. M., Laschi, F., Temussi, P., and Niccolai, N. (1999) *J. Biomol. NMR* **15**, 125–133
53. Esposito, G., Lesk, A. M., Molinari, H., Motta, A., Niccolai, N., and Pastore, A. (1992) *J. Mol. Biol.* **224**, 659–670
54. Yuan, T., Ouyang, H., and Vogel, H. J. (1999) *J. Biol. Chem.* **274**, 8411–8420
55. Bernini, A., Venditti, V., Spiga, O., Ciutti, A., Prischi, F., Consonni, R., Zetta, L., Arosio, I., Fusi, P., Guagliardi, A., and Niccolai, N. (2008) *Bio-phys. Chem.* **137**, 71–75

## Interdomain Flexibility in Full-length MMP-1

56. Spiga, O., Bernini, A., Scarselli, M., Ciutti, A., Bracci, L., Lozzi, L., Lelli, B., Di Maro, D., Calamandrei, D., and Niccolai, N. (2002) *FEBS Lett.* **511**, 33–35
57. Respondek, M., Madl, T., Göbl, C., Golser, R., and Zangger, K. (2007) *J. Am. Chem. Soc.* **129**, 5228–5234
58. Nanga, R. P., Brender, J. R., Xu, J., Veglia, G., and Ramamoorthy, A. (2008) *Biochemistry* **47**, 12689–12697
59. Jarvet, J., Danielsson, J., Damberg, P., Oleszczuk, M., and Graslund, A. (2007) *J. Biomol. NMR* **39**, 63–72
60. Arumugam, S., Hemme, C. L., Yoshida, N., Suzuki, K., Nagase, H., Berjanskii, M., Wu, B., and Van Doren, S. R. (1998) *Biochemistry* **37**, 9650–9657
61. Impagliazzo, A., and Ubbink, M. (2004) *J. Am. Chem. Soc.* **126**, 5658–5659
62. Wang, T. Z., Frederick, K. K., Igumenova, T. I., Wand, A. J., and Zuiderweg, E. R. P. (2005) *J. Am. Chem. Soc.* **127**, 828–829
63. Van den Steen, P. E., Van Aelst, I., Hvidberg, V., Piccard, H., Fiten, P., Jacobsen, C., Moestrup, S. K., Fry, S., Royle, L., Wormald, M. R., Wallis, R., Rudd, P. M., Dwek, R. A., and Opdenakker, G. (2006) *J. Biol. Chem.* **281**, 18626–18637
64. Saffarian, S., Collier, I. E., Marmer, B. L., Elson, E. L., and Goldberg, G. (2004) *Science* **306**, 108–111
65. Overall, C. M., and Butler, G. S. (2007) *Structure (Lond.)* **15**, 1159–1161
66. Ottl, J., Gabriel, D., Murphy, G., Knauper, V., Tominaga, Y., Nagase, H., Kroger, M., Tschesche, H., Bode, W., and Moroder, L. (2000) *Chem. Biol.* **7**, 119–132

## Characterisation of the MMP-12–Elastin Adduct

Ivano Bertini,<sup>\*,[a, b]</sup> Marco Fragai,<sup>[a, c]</sup> Claudio Luchinat,<sup>[a, c]</sup> Maxime Melikian,<sup>[a]</sup> and Chiara Venturi<sup>[a]</sup>

*Dedicated to the Centenary of the Italian Chemical Society*

MMP-12 (matrix metalloproteinase 12) is an important protein of the MMP family.<sup>[1–4]</sup> Its substrate is elastin,<sup>[5–8]</sup> which is composed of a cross-linked insoluble network of polypeptide chains of tropoelastin of MW 65 kDa each.<sup>[9]</sup> Insoluble elastin is responsible for keeping some extracellular connective tissues elastic.

The full-length MMP-12 is made up by a catalytic (CAT) domain and a hemopexin-like (HPX) domain. The two domains have relatively large conformational freedom,<sup>[10]</sup> as also recently found in MMP-1<sup>[11]</sup> and MMP-9.<sup>[12]</sup> The role of the HPX domain is not certain, as the CAT domain is always active by itself,<sup>[13]</sup> even if it may have a smaller turnover for some substrates.<sup>[14]</sup>

Experimental structures and models, representative of the binding mode of peptide substrates on protein active site are already present in the Protein Data Bank. However, the interaction of native elastin with full length MMP-12 and its isolated domains had not been structurally investigated so far. Here we report a NMR study of MMP-12 with elastin. MMP-12, with the catalytic zinc replaced by cadmium<sup>[10,15]</sup> and with the catalytically competent Glu219 replaced by alanine, is inactive<sup>[16]</sup> and allows the investigation of its interaction with elastin without cleavage of the substrate. Insoluble elastin can be made soluble by limited proteolysis<sup>[17]</sup> that

provides a mixture of cross-linked polymers of several tens of kDa or more. The soluble mixture from bovine neck ligament (elastin hereafter) is the closest viable model to native elastin. This protein shares 76% identity (excluding long insertions/deletions) with the human ortholog. The assignment of the <sup>13</sup>C-, <sup>2</sup>H- and <sup>15</sup>N-labelled full-length MMP-12 protein is available.<sup>[10]</sup>

In the presence of elastin some <sup>1</sup>H-<sup>15</sup>N HSQC signals are shifted, as reported in Figure 1A. Shifted signals are observed both in the CAT and HPX domains; in the former these are more numerous and a decrease in intensity of several signals is also observed. The spectra of the isolated CAT domain in the presence of elastin undergo similar shifts and intensity decreases as those observed for the full-length protein (Figure 1B) and the isolated HPX domain also maintains similar shifts (Figure 1C). Comparison of Figures 1B and 1C with 1A shows that both interactions are clearly operative separately, although an overall reinforcement of the interaction of the CAT domain in the full-length protein may be noted, possibly brought about by a cooperative behaviour of the HPX domain. A comparison of the chemical shift changes with their signs in the full-length protein and in the isolated domains is reported in Supporting Information (see Figure S1). The pattern of residues shifted upon addition of elastin is largely similar in the full length protein and in the isolated domains, but is more pronounced in the former (Figure S2). The overall pattern is consistent with a simultaneous interaction of both domains with the elastin polymer, also in view of the recently reported conformational flexibility of the two domains.<sup>[10]</sup> While this finding suggests that the HPX domain may have a role in elastin recognition, it is clear that the isolated CAT domain is essentially able to recognise its natural substrate by itself.

By analyzing in more detail the pattern of shifted NH nuclei in the isolated CAT domain, it is immediately apparent that the affected nuclei all belong to the active crevice and to the internal basal layer of the catalytic domain of MMP-12 (Figure 2).<sup>[10]</sup> These results demonstrate the occur-

[a] Prof. I. Bertini, Dr. M. Fragai, Prof. C. Luchinat, Dr. M. Melikian, Dr. C. Venturi  
Magnetic Resonance Center (CERM), University of Florence  
Via L. Sacconi 6, 50019 Sesto Fiorentino (Italy)  
Fax: (+39)0554574271  
E-mail: bertini@cerm.unifi.it

[b] Prof. I. Bertini  
Department of Chemistry, University of Florence  
Via della Lastruccia 3, 50019 Sesto Fiorentino (Italy)

[c] Dr. M. Fragai, Prof. C. Luchinat  
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://dx.doi.org/10.1002/chem.200901009>.

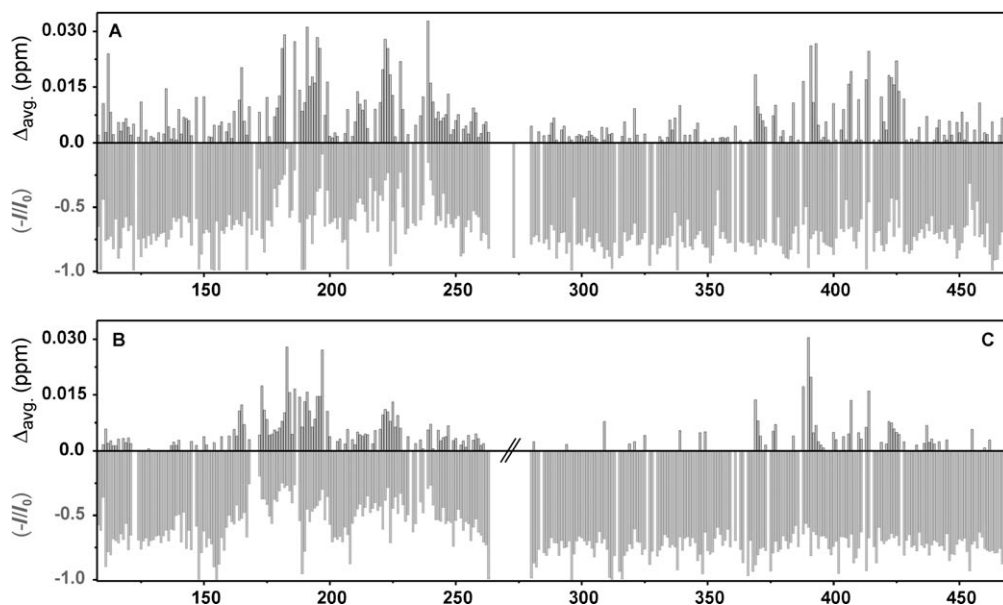


Figure 1. Plot of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC chemical shifts ( $\{[(\delta_{\text{H}})^2 + (\delta_{\text{N}}/5)^2]/2\}^{1/2}$ )<sup>[18]</sup> and normalised intensity changes per residue observed in 0.35 mM full-length MMP-12 (A) and 0.30 mM of its isolated CAT (B) and HPX (C) domains upon addition of  $1.4 \text{ mg mL}^{-1}$  elastin.

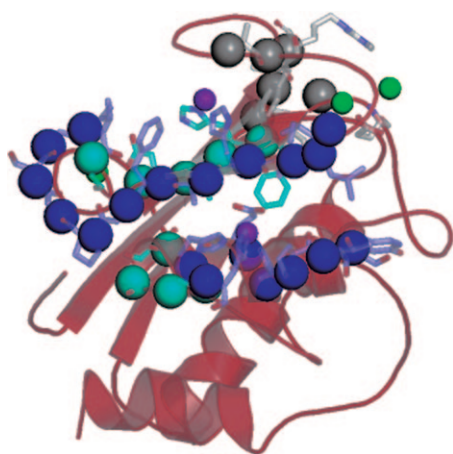


Figure 2. Residues of the CAT domain of MMP-12 experiencing substantial NH shifts/intensity changes upon addition of elastin, highlighted on the atomic resolution structure<sup>[25]</sup> of the protein. The dark blue spheres indicate the NH groups of the residues at the forefront of the interaction with elastin, while the sky blue spheres indicate nonsolvent-exposed residues in the rear  $\beta$ -strand that are indirectly affected by the interaction with elastin through the hydrogen-bonding network. The grey spheres represent the NH groups of the residues experiencing a smaller interaction which is not directly connected with the main one and is not analyzed further. The side chains of the residues involved are shown as sticks of the corresponding colors. The magenta sphere represent zinc ions, the green ones the calcium ions.

rence of a specific localised interaction of elastin with the catalytically competent site of MMP-12 and, therefore, indicate that the only relevant interaction(s) indeed concerns recognition of the elastin cleavage site(s). This is a nice result, although in line with expectations. Of course, the detected specificity is only on the MMP-12 side and not necessarily on the elastin side. Actually, there are as many as 36

cleavage sites in insoluble elastin, detected by proteolysis and mass spectrometry,<sup>[7]</sup> and many more (86 sites) in tropoelastin,<sup>[19]</sup> on account of the increased number of accessible cleavage sites in the absence of cross-linking.<sup>[7,19]</sup> We know<sup>[7]</sup> that most of the cleavage sites in elastin occur at G-L (30 sites) and some at G-V (11 sites) and A-L (11 sites). Selected sequences that correspond to cleavage sites that appear in more than one peptide fragment are shown in Table 1 for both bovine and human elastin. We calculated the structures of the adducts between CAT MMP-12 and each polypeptide in Table 1 by positioning the CO to be

Table 1. Selected bovine and human elastin sequences (P=4-hydroxyproline) that correspond to cleavage sites appearing in more than one peptide fragment.<sup>[7]</sup> All peptides were tested using the program Haddock<sup>[21]</sup> for their interaction with the active residues in the CAT domain (dark blue spheres in Figure 2), starting from 6-AA long fragments centred at the G-L, A-L or G-V cleavage sites (in bold) up to a length of 15 residues. The length of each fragment is that showing optimal Haddock interaction parameters in each case, as judged by the best 40 structures for each run.

1	bovine(43–53) human (44–53)	FPGAGLG <b>GLGV</b> PGAGLGALGG
2	bovine(64–72) human (64–72)	KpGV <b>GLVG</b> KpVp <b>GGLAG</b>
3	bovine(554–563) human (574–583)	QFRAA <b>AGLPA</b> QLRAA <b>AGLGA</b>
4	bovine(573–581) human (593–601)	GVPGL <b>GVGA</b> GVPGL <b>GVGA</b>
5	bovine(628–635) human (681–688)	GIp <b>GGVAG</b> GIp <b>GGVVG</b>
6	bovine(649–657) human (703–711)	KAAQF <b>GLGG</b> KAAQF <b>GLVG</b>
7	bovine(659–668) human (715–725)	GGLGV <b>GGLGA</b> LGLGV <b>GGLGV</b>
8	bovine(668–677) human (724–733)	AVPGA <b>VLGG</b> GVPGV <b>GGLGG</b>

cleaved (G or A) close to the metal-coordinated catalytic water,<sup>[16,20]</sup> the hydrophobic leucine or valine side chains in the S<sub>1</sub>' pocket and by assuming that the MMP-12 residues interacting with the peptide are those that experience shifts and intensity reduction on the protein surface (dark blue spheres in Figure 2). The program Haddock was used.<sup>[21]</sup> We found that a sequence of 4–7 amino acids (AA) before G-L (or G-V or A-L) and two after them are enough to justify the NMR shifts. This finding is absolutely consistent with the mechanism previously proposed,<sup>[16]</sup> and with the available structures and models representative of the binding mode of the peptide substrates on the protein active site.<sup>[20,22–24]</sup>

A representative example of a CAT–peptide complex is shown in Figure 3. It can be noted that the peptide nicely fits within what is known as the active site crevice and forms

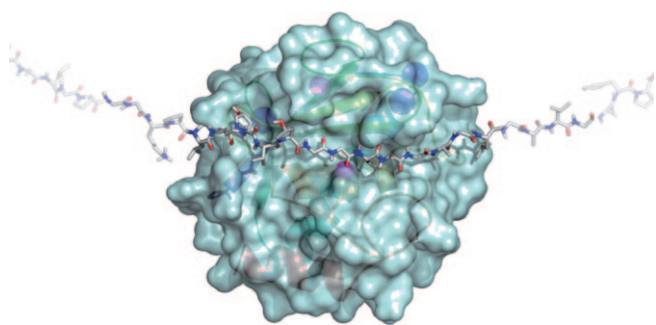


Figure 3. Complex between the CAT domain of MMP-12 and an elastin fragment incorporating the 9-AA peptide **2** (corresponding to residues 64–72 in human elastin) in Table 1. Additional residues on both sides of the 9-AA peptide have been added to recall the macromolecular nature of the substrate.

a number of hydrogen-bonding interactions with it (Table S1; see Supporting Information). There is a striking similarity between the present peptide binding mode and the X-ray-deduced structure of a 6-AA collagen-derived peptide and its cleaved fragments bound to MMP-12 CAT domain, strongly corroborating the present results.<sup>[20]</sup> The peptide in Figure 3 is prolonged by a few residues in both directions to highlight the fact that the detected interaction is not with a small peptide but with macromolecular elastin. The heterogeneous composition of the elastin sample and the cross-linked structure of the polymer prevent the identification of longer peptides possibly able to simultaneously bind to both the catalytic and the hemopexin domains. On the other hand, the elastin parts responsible for the interaction with the HPX domain are not necessarily close in sequence to those interacting with the CAT domain and may even belong to cross-linked stretches originated from different tropoelastin monomers.

In conclusion, light has been shed on the recognition by MMP-12 of its natural substrate elastin. The interaction occurs at both the CAT and HPX sites, the former being stronger and largely maintained in the isolated CAT domain, consistent with the reported ability of the isolated

CAT domain to retain substantial catalytic activity. The pattern of affected residues on the CAT domain very clearly points to an extended interaction within the active site crevice, involving 8–11 elastin residues. Although the structural features of cross-linked elastin are largely unknown, it is interesting to note that such extended stretches of residues, as they appear from the present work to be necessary for a productive interaction, are likely to be available for interaction with MMP-12, because they are generally far away in sequence from the lysines involved in cross-linking, fully consistent with the observation that no cleavage sites are observed in proximity of lysines.

## Experimental Section

Full-length MMP-12, its catalytic domain and the hemopexin-like domain were expressed and purified as described previously.<sup>[10,16]</sup> The replacement of the catalytic zinc with the inactive cadmium ion was performed by exhaustive dialysis against a buffer containing 20 mM Tris pH 7.2, 10 mM CaCl<sub>2</sub>, 0.15 M NaCl, 0.2 M acetohydroxamic acid (AHA) and 0.3 mM of CdCl<sub>2</sub>. The acetohydroxamic acid was removed with the last dialysis. Soluble elastin from bovine neck ligament was purchased from Sigma–Aldrich. Lyophilised elastin was dissolved in the final buffer (20 mM Tris pH 7.2, 10 mM CaCl<sub>2</sub>, 0.15 M NaCl and 0.3 mM of CdCl<sub>2</sub>) and extensively dialyzed before the addition to the protein solution. Soluble elastin (0.7 mg) was added to 0.5 mL of full length MMP-12 (0.35 mM). The same amount of soluble elastin was added to 0.5 mL of isolated catalytic and hemopexin domains at a concentration of 0.3 mM. The final concentration of soluble elastin in solution, calculated as tropoelastin (MW ~65 KDa), was estimated to be  $2.2 \times 10^{-2}$  mM. However, taking into account the presence of several cleavage sites in the substrate (up to 86 in tropoelastin and up to 36 in insoluble elastin<sup>[7,19]</sup>) the concentration of binding sites for the catalytic domain in the samples may range from 0.8 up to 1.9 mM, that is, in large stoichiometric excess. The heterogeneous composition of soluble elastin and the presence of several binding sites, plausibly with different affinities, prevent the calculation of any dissociation constant.

## Acknowledgements

This work was supported by EC (Projects: MEST-CT-2004-504391, SFMET n° 201640, SPINE2-COMPLEXES n° 031220 and Nano4Drugs n° LSHB-CT-2005-019102), by MIUR (PRIN 2005, Prot. N. 2005039878, Prot. RBLA032M7, Prot. RBIP06 LSS2) and Ente Cassa di Risparmio di Firenze.

**Keywords:** enzyme catalysis • metalloenzymes • NMR spectroscopy • peptides • reaction mechanisms

- [1] S. D. Shapiro, *Thromb. Haemostasis* **1999**, 82, 846–849.
- [2] J. Liang, E. Liu, Y. Yu, S. Kitajima, T. Koike, Y. Jin, M. Morimoto, K. Hatakeyama, Y. Asada, T. Watanabe, Y. Sasaguri, S. Watanabe, J. Fan, *Circulation* **2006**, 113, 1993–2001.
- [3] C. M. Overall, *Mol. Biotechnol.* **2002**, 22, 51–86.
- [4] L. Aureli, M. Gioia, I. Cerbara, S. Monaco, G. F. Fasciglione, S. Marini, P. Ascenzi, A. Topai, M. Coletta, *Curr. Med. Chem.* **2008**, 15, 2192–2222.
- [5] R. P. Mecham, T. J. Broekelmann, C. J. Fliszar, S. D. Shapiro, H. G. Welgus, R. M. Senior, *J. Biol. Chem.* **1997**, 272, 18071–18076.



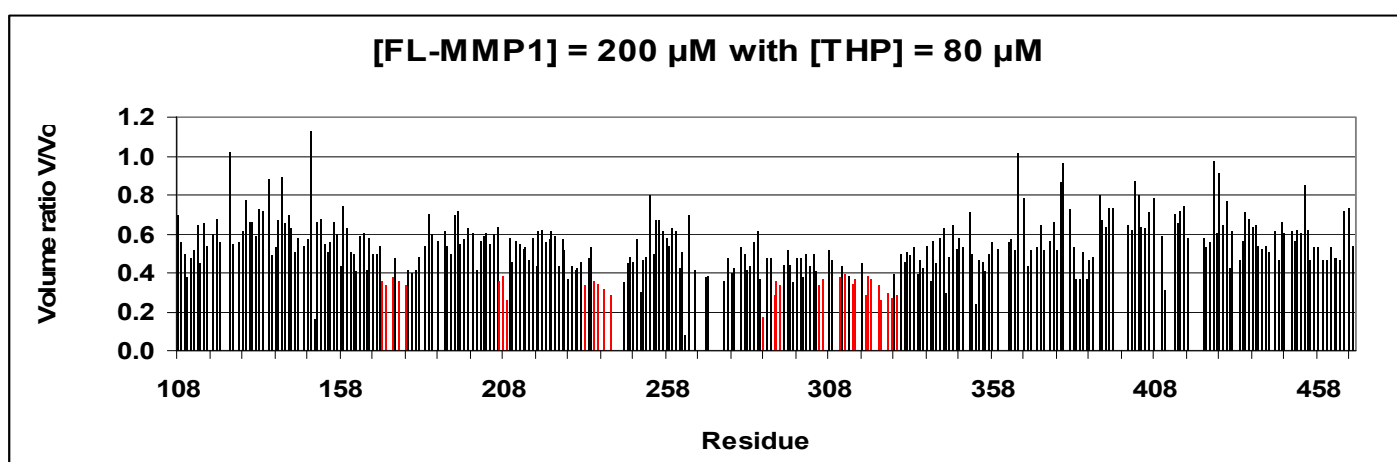
- [6] T. J. Gronski, R. L. Martin, D. K. Kobayashi, B. C. Walsh, M. C. Holman, M. Huber, H. VanWart, S. D. Shapiro, *J. Biol. Chem.* **1997**, *272*, 12189–12194.
- [7] S. Taddese, A. S. Weiss, R. H. Neubert, C. E. Schmelzer, *Matrix Biol.* **2008**, *27*, 420–428.
- [8] S. M. Mithieux, A. S. Weiss, *Adv. Protein Chem.* **2005**, *70*, 437–461.
- [9] S. G. Wise, A. S. Weiss, *Int. J. Biochem. Cell Biol.* **2009**, *41*, 494–497.
- [10] I. Bertini, V. Calderone, M. Fragai, R. Jaiswal, C. Luchinat, M. Melikian, E. Mylonas, D. Svergun, *J. Am. Chem. Soc.* **2008**, *130*, 7011–7021.
- [11] I. Bertini, M. Fragai, C. Luchinat, M. Melikian, E. Mylonas, N. Sarti, D. Svergun, *J. Biol. Chem.* **2009**, *284*, 12821–12828.
- [12] G. Rosenblum, P. E. Van den Steen, S. R. Cohen, J. G. Grossmann, J. Frenkel, R. Sertchook, N. Slack, R. W. Strange, G. Opendakker, I. Sagi, *Structure* **2007**, *15*, 1227–1236.
- [13] J. M. Shipley, G. A. Doyle, C. J. Fliszar, Q. Z. Ye, L. L. Johnson, S. D. Shapiro, H. G. Welgus, R. M. Senior, *J. Biol. Chem.* **1996**, *271*, 4335–4341.
- [14] R. Bhaskaran, M. O. Palmier, J. L. Lauer-Fields, G. B. Fields, S. R. Van Doren, *J. Biol. Chem.* **2008**.
- [15] I. Bertini, M. Fragai, Y.-M. Lee, C. Luchinat, B. Terni, *Angew. Chem.* **2004**, *116*, 2304–2306; *Angew. Chem. Int. Ed.* **2004**, *43*, 2254–2256.
- [16] I. Bertini, V. Calderone, M. Fragai, C. Luchinat, M. Maletta, *Angew. Chem.* **2006**, *118*, 8120–8123; *Angew. Chem. Int. Ed.* **2006**, *45*, 7952–7955.
- [17] S. Keller, I. Mandl, *Biochem. Med.* **1971**, *5*, 342–347.
- [18] S. Grzesiek, A. Bax, G. M. Clore, A. M. Gronenborn, J. S. Hu, J. Kaufman, I. Palmer, S. J. Stahl, P. T. Wingfield, *Nat. Struct. Biol.* **1996**, *3*, 340–345.
- [19] S. Taddese, A. S. Weiss, G. Jahreis, R. H. Neubert, C. E. Schmelzer, *Matrix Biol.* **2009**, *28*, 84–91.
- [20] I. Bertini, V. Calderone, M. Fragai, C. Luchinat, S. Mangani, B. Terni, *Angew. Chem.* **2003**, *115*, 2777–2780; *Angew. Chem. Int. Ed.* **2003**, *42*, 2673–2676.
- [21] S. J. de Vries, A. D. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar, A. M. Bonvin, *Proteins Struct. Funct. Bioinf.* **2007**, *69*, 726–733.
- [22] B. Lovejoy, A. M. Hassell, M. A. Luther, D. Weigl, S. R. Jordan, *Biochemistry* **1994**, *33*, 8207–8217.
- [23] F. Grams, V. Dive, A. Yiotakis, I. Yiallourous, S. Vassiliou, R. Zwilling, W. Bode, W. Stocker, *Nat. Struct. Biol.* **1996**, *3*, 671–675.
- [24] R. Lang, A. Kocourek, M. Braun, H. Tschesche, R. Huber, W. Bode, K. Maskos, *J. Mol. Biol.* **2001**, *312*, 731–742.
- [25] I. Bertini, V. Calderone, M. Cosenza, M. Fragai, Y.-M. Lee, C. Luchinat, S. Mangani, B. Terni, P. Turano, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5334–5339.

Received: April 16, 2009  
Published online: July 16, 2009

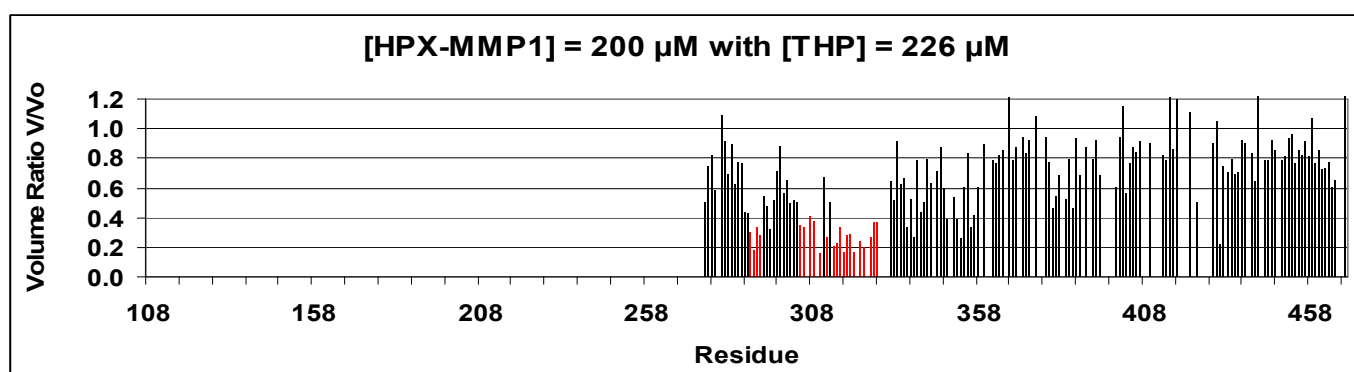
### 3.4 Interaction between MMP1 and a collagen peptide model

#### 3.4.1 Interaction between FL-MMP1 and triple helical peptide (THP)

Titration experiment were performed using a triple labelled FL-MMP1 Cadmium sample containing a E219A mutation by adding volumes of concentrated THP. After addition of THP, TROSY spectra obtained resulted in a general intensity decrease of MMP1 signals (figure 11). However, the main decrease was observed for the HPX domain in the region located between residues I290-A330. This region corresponds to a group of three beta-sheets and one alpha-helix located on the surface constituting one of the four-blade beta-propeller of the HPX motif. Residues from the CAT domain also experienced some intensity decrease : in particular, in the segments N171-N180 and F207-E209, for the residues H228 and G233 and in the segment P238-T241. These residues are located at the protein surface on the side of the active site.



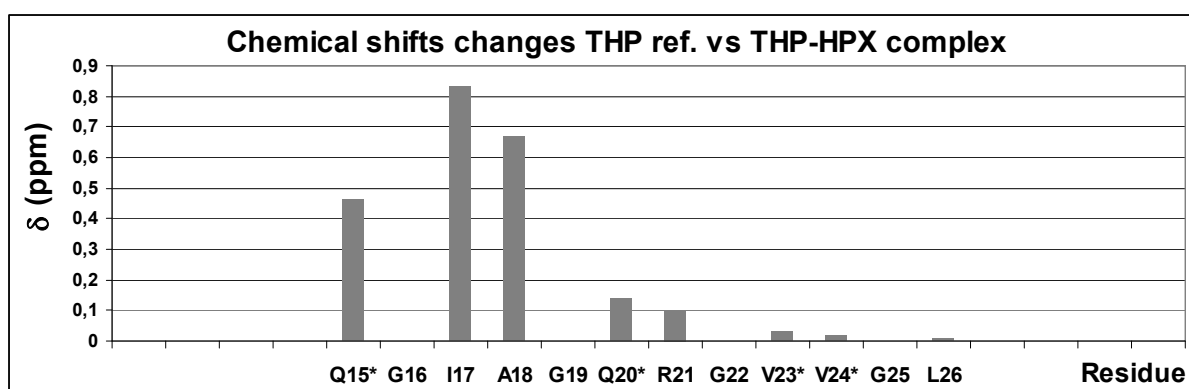
**Figure 11.** Peaks intensity decrease of FL-MMP1 by adding unlabelled THP for HN signals on TROSY spectra acquired at 310K on 900 MHz spectrometer.



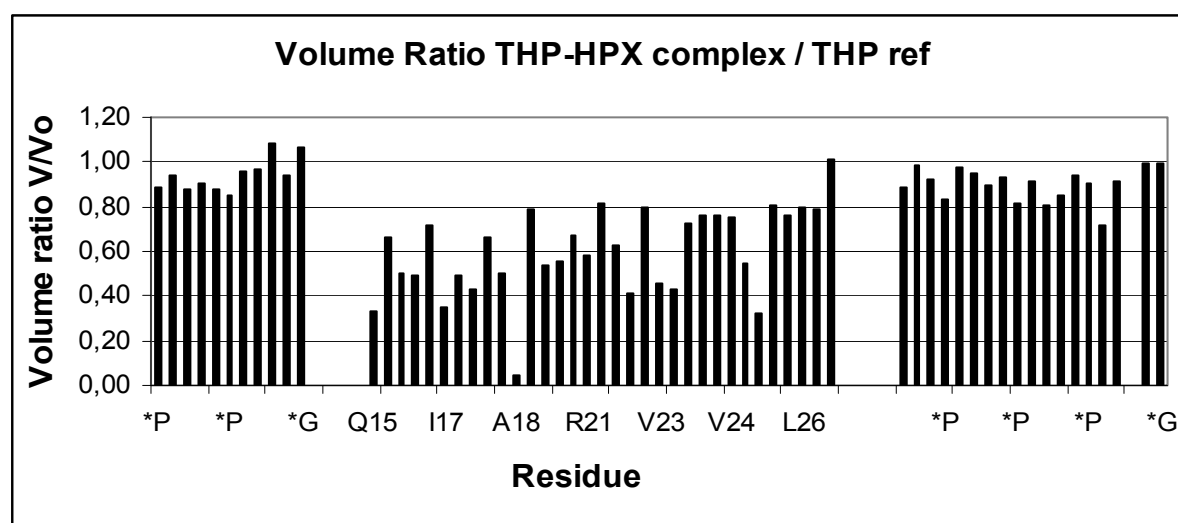
**Figure 12.** Peaks intensity decrease of HPX-MMP1 by adding unlabelled THP for HN signals on TROSY spectra acquired at 310K on 900 MHz spectrometer.

### 3.4.2 Interaction between HPX-MMP1 and THP

In order to confirm the interaction between HPX domain and the THP, a similar titration experiment was performed using isolated HPX of MMP1 with THP (figure 12). Intensity decrease observed in the isolated HPX domain corresponds to the same regions as observed in HPX domain inside the FL protein from the previous titration (figures 11 and 12). By looking into details, this HPX domain region is poorly conserved among the MMP family. For example, by fitting together HPX domains from MMP12 and MMP1 structures, it appears the main MMP1 fragment M305-R313 involved this interaction has a completely different backbone trajectory respect with the corresponding one in MMP12. This region may be important for MMP1 collagenase specificity. Moreover, this HPX domain area is partially covered by the pro-domain in the pro-MMP1 X-ray structure. Then, MMP1 pro-domain function could be not only to lock the CAT domain active site but also to prevent the HPX domain binding to collagen.



**Figure 13.** Chemical shift changes observed on THP by adding unlabelled HPX-MMP1 for proton signals on TOCSY spectra acquired at 310K on 700 MHz spectrometer. An asterisk\* indicates repeated amino acids in the central region which haven't been assigned sequentially.



**Figure 14.** Peaks intensity decrease on THP by adding unlabelled HPX-MMP1 for proton signals on TROSY spectra acquired at 310K on 700 MHz spectrometer. An asterisk\* indicates repeated gly, pro and hyp in the lateral part of the peptide.

Complementary 2D HH TOCSY experiments performed on THP in absence and in presence of HPX allowed visualising important chemical shift differences on THP proton spin (figure 13). Although THP is unlabelled and contains several gly, pro and hyp residues which cannot be assigned sequentially, other residues can be identified. Else, chemical shifts changes were observed for ile, ala and arg (one residue per chain) and for gln and val (two residues per chain). All these residues are concentrated in the central part of the (GPHyp)<sub>4</sub>-**GPQ-GIA-GQR-GVV**-GLP-(GPHyp)<sub>4</sub> peptide. In addition, a broad cluster of signals corresponding to prolines and hydroxyprolines spin systems didn't experience any chemical shift changes after addition of the HPX domain.

Several volumes of THP cross-peak signals well resolved on the TOCSY spectra were integrated on the THP alone reference spectra and in the spectra of the complex HPX-THP (figure 14). While for some unassigned spin systems identified as prolines, hydroxyprolines and glycines, the intensity ratio remained close to 1, for the other amino-acid located in the central region of the peptide, intensity ratio decreased between 0.8 and 0.3. The strongest diminutions were measured for I17 and A18.

### 3.4.3 Docking calculation for hemopexin-like domain of MMP1 and THP from chemical shift perturbation using Haddock program

The docking strategy has consisted in first considering separately HPX domain and CAT domain of MMP1 to interact with THP.

For the HPX domain, active residues were selected from both chemical shift perturbation and solvent accessibility. The threshold for peaks intensity decrease of HPX domain was defined at 35% ratio  $I/I_0$  (figure 12). This value corresponds to the average value minus the standard deviation value. HPX domain residue accessibility was calculated from the MMP1 X-ray structure (accession code 2CTL). For the residue selection, threshold for solvent accessibility given by  $N_{\text{access}}$  was set at 50%.

Then, the following residues were selected from HPX domain :

R291, G292, N306, Y309, E311, E313, N315, F316, S318, V319 and F320.

For the THP, residues were selected in the central region on the basis of experimental data from 2D-TOCSY experiments (figure 13 and 14).

Q15, I17, A18, Q20, R21, V23, V24,  
Q54, I56, A57, Q59, R60, V62, V63,  
Q93, I95, A96, Q98, R99, V101, V102

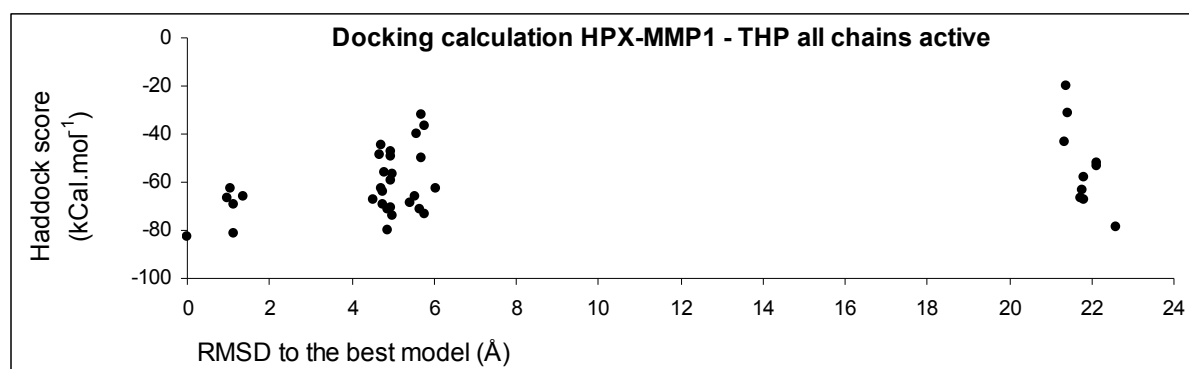


Figure 15. Haddock docking calculation HPX with THP with all chains selected

The results from this run of calculation provided three clusters located respectively at 1 Å, 5 Å and 21 Å from the best Haddock score :

$$\text{Haddock Score} = E_{\text{vdw}} + 0.2 E_{\text{elec}} + 0.1 E_{\text{viol}} + E_{\text{desolv}}$$

In the cluster located at 21 Å from the best model (figure 15), the HPX domain orientation is not compatible with the CAT domain orientation suitable for enzymatic cleavage on collagen. Then, this solution was discarded. From the structures on the two remaining clusters the contact area observed between HPX domain and THP is concentrated around two of the three chains forming the triple helix. From this result, the third peptidic chain away from the HPX domain in the two main clusters was discarded for the next calculation.

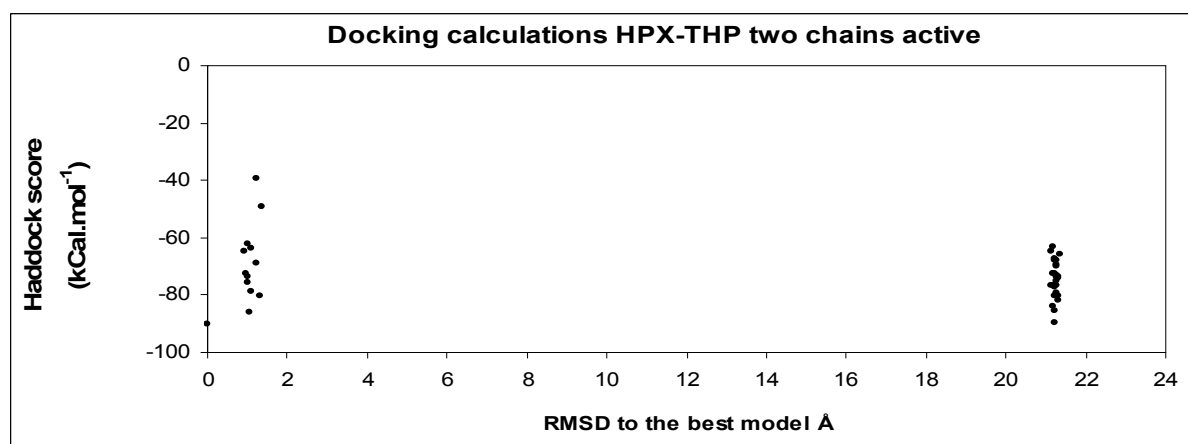
In the following step, calculation was run using only two chains of the triple helical peptide.

Q54, I56, A57, Q59, R60, V62, V63,  
Q93, I95, A96, Q98, R99, V101, V102

The same active residues as before were selected on HPX domain.

R291, G292, N306, Y309, E311, E313, N315, F316, S318, V319 and F320.

This time, only fourteen residues are selected from the peptide which is more likely to interact with eleven residues from the HPX domain.



**Figure 16.** Haddock docking calculation HPX with THP with two chains selected

Here, calculation resulted in only two clusters (figure 16). HPX domain binds the THP at the same position in the two clusters but with two opposite orientations. Compared with the first calculation, the two clusters are more homogeneous. Here, the cluster closer to the best calculated structure is not compatible with the CAT domain orientation suitable for cleavage, and then this solution was also discarded. In the other hand, the cluster at 21 Å of distance to the best structure has a HPX domain orientation compatible with the CAT domain position suitable for cleavage. Moreover, the latter cluster contains more structures (27 versus 13) and the two clusters have a comparable Haddock score. Then, the main cluster for the HPX-THP complex was selected to monitor the FL-MMP1 docking to THP.

### 3.4.4 Docking calculation on catalytic domain of MMP1 and THP (Fields) using Haddock program

Prior to describe the interaction between FL-MMP1 and THP, a model for collagen peptide insertion inside the CAT domain active site is required.

On the native collagen, the cleavage site corresponds to a GLY-ILE on  $\alpha 1$  chain. However, no direct information about the interaction between the catalytic domain of MMP1 and THP has been reported to date. From NMR titration of FL-MMP1 with THP (figure 11), some CAT-MMP1 residues were selected from their intensity decrease and solvent accessibility. Additionally, relying on the CAT-MMP1 homology structure with the CAT-MMP12, the latter selection was completed by adding residues experiencing chemical shift changes from titration between MMP12 and elastin. Indeed most residues selected from CAT-MMP12 are located in the same area as the residues from CAT-MMP1 even if the positions are lightly different.

A first Haddock calculation for CAT-THP complex was run using the active residue from both CAT-MMP1 and CAT-MMP12 using CAT-MMP1 structure (from 2CLT) for the docking. Moreover, distance constraints were added to maintain the coordination between the catalytic zinc metal ion and the three histidines 218, 222, 228. Else, according the catalytic mechanism, other distance constraints were added to position properly the catalytic zinc ion close to the G-I cleavage site on collagen peptide. These distances were previously observed in a MMP12-peptidic ligand X-ray structure.

From the previous calculation run with the HPX domain, only one chain from the triple helix remains accessible for the CAT domain. Then, residues selected on THP correspond to a fragment around the G-I cleavage site on the accessible chain for CAT domain. Docking calculation results are shown in figure 17.

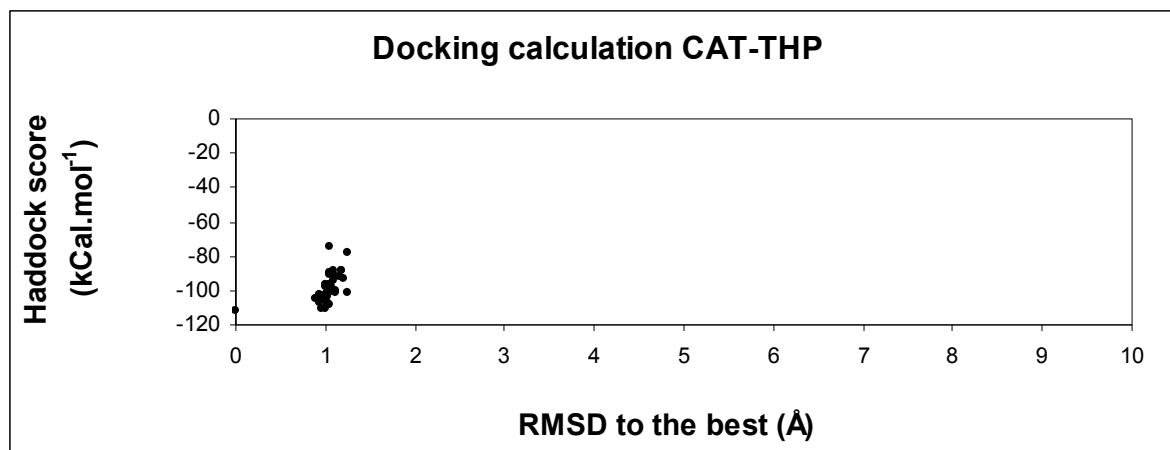


Figure 17. Haddock docking calculation CAT with THP with one chain selected

Active residues on CAT domain

\_ From FL-MMP1 titration with THP: 171, 178, 180, 207, 208, 209, 228, 233, 239 and 241  
\_ From CAT-MMP12 interaction with ligands : 180, 181, 182, 183, 184, 185, 186, 187, 218, 219, 222, 228, 238, 239 and 240

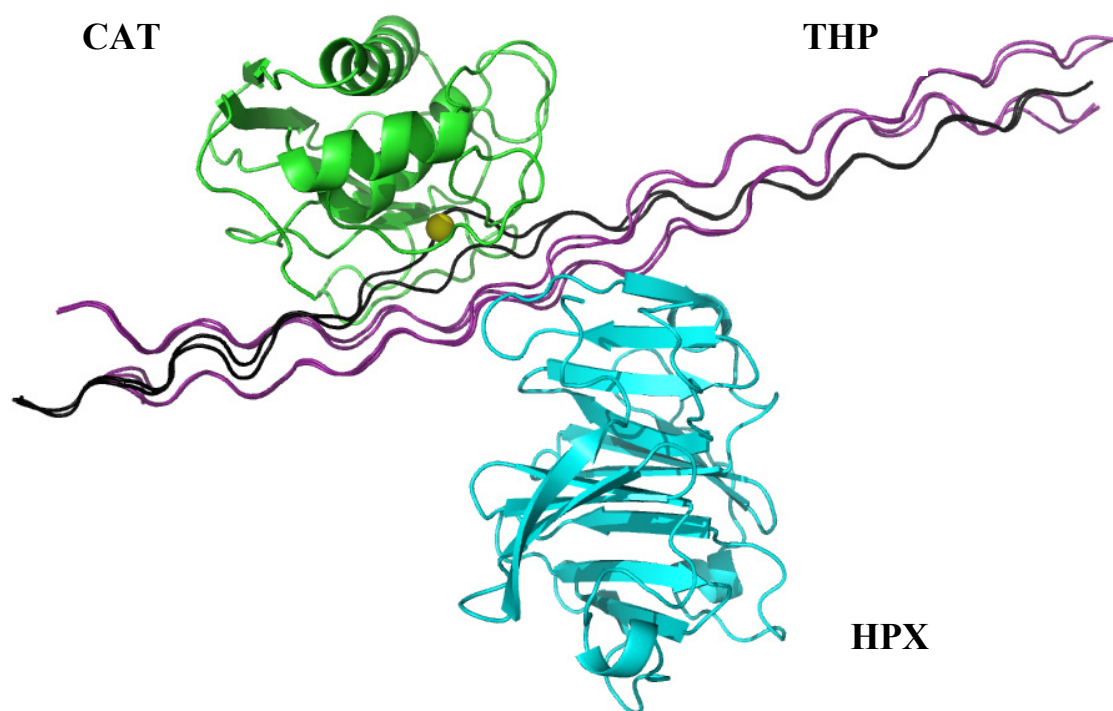
Active residues on THP

\_ From THP around cleavage site : G13, P14, Q15, G16, I17, A18, G19, Q20, R21

Calculations resulted in a unique cluster where the CAT domain reached an appropriate position for cleavage. Again, this CAT-THP interaction model obtained can be reported to the final calculation within the FL-protein with the THP.

### 3.4.5 Fitting of MMP1 single domains to the THP from calculated complexes

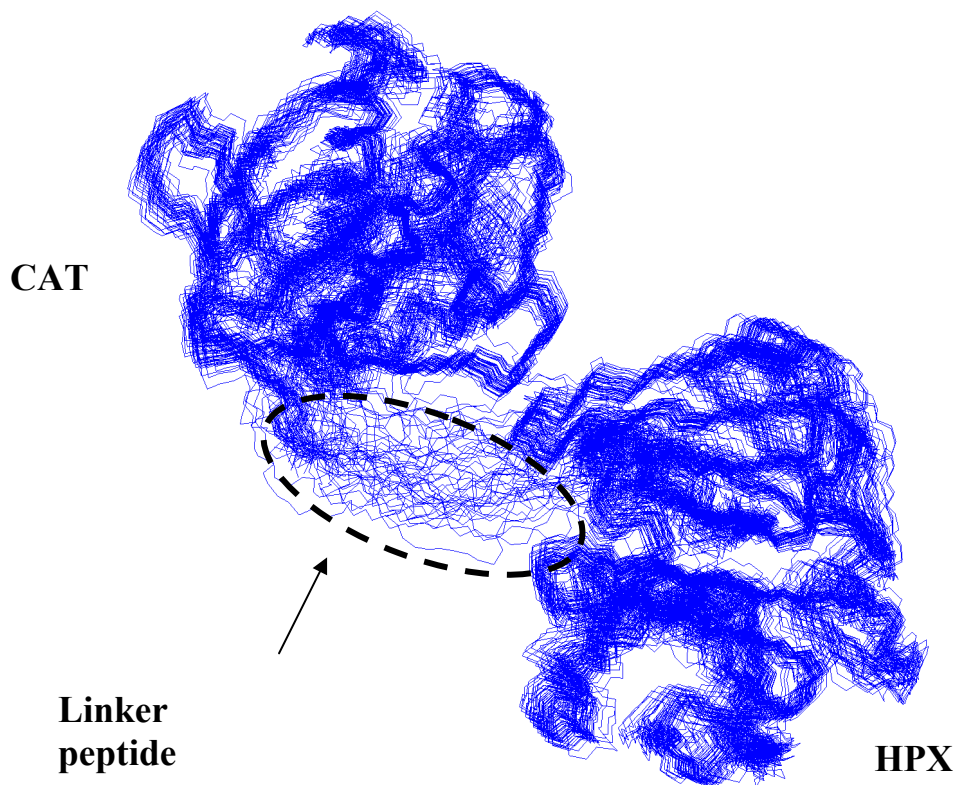
Once calculated the model structures for CAT-THP and HPX-THP separately, both complexes were fit on THP to define the single domain positions in a new model for a compatible structure with FL-MMP1. Here “simultaneous” interaction of HPX and CAT domain of FL-MMP1 with THP are required. Importantly, linker peptide extremities from each domain of MMP1 are located at a compatible distance whereas for the discarded HPX-THP clusters, virtual distance between extremities of the linker was not compatible with the real linker length.



**Figure 18.** Fitting of the two appropriate models HPX-THP and CAT-THP on the THP using molmol. Figure was displayed using pymol.

### 3.4.6 FL-MMP1 Structure from THP-fixed single domains (HPX and CAT) using Modeller program

Starting from the previous fitting (figure 18), the two reciprocally oriented domains were used as a scaffold in a calculation by using modeller. FL-MMP1 linker peptide was generated to provide the family of structures represented below (figure 19). By looking at the ensemble, numerous conformations for the linker peptide between the two domains were obtained. However considering the significant distance between the linker extremities, the preferential trajectory is relatively straight. From calculations, the structures were sorted on the basis of an energy term.



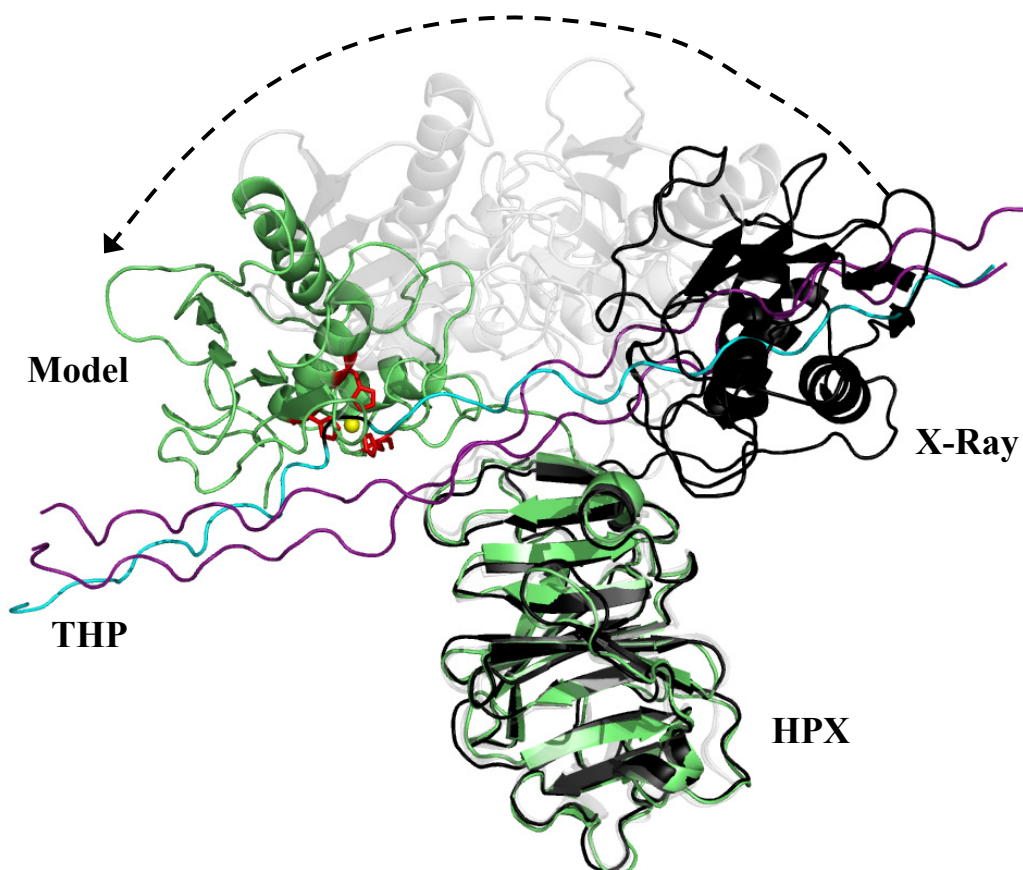
**Figure 19.** Modeller calculation of FL-MMP1 from FIT CAT-THP – HPX-THP. Family of 49 structures where the linker peptide was generated keeping the two domains in their initial positions from the previous fitting.



### 3.4.7 Docking calculation by Haddock program using chemical shift perturbation on Modeller structure of FL-MMP1 and THP (Fields)

A suitable FL-MMP1 structure was extracted from Modeller ensemble, and Haddock calculations were run combining all the constraints from the single domains docking. From 50 structures, a unique cluster was obtained with an average RMSD of 0.5 Å, Van der Waals energy = -123.5 +/- 3.6 kCal.mol<sup>-1</sup>, electrostatic energy = -419.8 +/- 31.9 kCal.mol<sup>-1</sup>, desolvation energy = -5.1 +/- 6.6 kCal.mol<sup>-1</sup>, restraints violation energy = 125.2 +/- 24.85 kCal.mol<sup>-1</sup> and a buried surface area of 3143.8 +/- 38.3 Å<sup>2</sup>.

The final model obtained is shown here (figure 20). HPX-MMP1 interacts with two chains from THP and the CAT-MMP1 interacting with the released chain of the partially unwound THP. Final model obtained was superimposed with the X-Ray structure fitting on the HPX domains. In the final model, CAT domain reaches the cleavage site by a rotation of the linker peptide around the THP.



**Figure 20.** Interaction model of the FL-MMP1 THP collagen-like complex

**CAT active residues :** N171, G178, N180, L181, A182, H183, A184 , F185, Q186, F207, T208, E209, H218, H222, H228, G233, P238, S239, Y240, T241

**Fields peptide :** P14, Q15, G16, I17, A18, G19, Q20, R21 Q54, I56, A57, Q59, R60, Q98, R99, V101 e V102

**HPX active residues :** R291, G292, N306, Y309, E311, E313, N315, F316, S318, V319, F320

**Forced distances :** metal ion position, hbonds CAT-THP and HPX-THP and distances from catalytic site to cleavage site.

## **4. Conclusion**

## *Mobility of Matrix metalloproteinases MMP12 and MMP1*

During this research the question of inter-domain flexibility in solution for MMPs bearing short linkers has been addressed by NMR.<sup>1,2</sup> In both MMP1 and MMP12 the <sup>1</sup>H-<sup>15</sup>N HSQC or TROSY spectra of the FL construct is largely the superposition of the related catalytic and hemopexin domain spectra, consistent with the absence of an extensive contact between the two domains. The absence of a stable contact between the HPX and CAT domain in MMP1 has been also observed by monitoring the effect of a soluble paramagnetic probe on the NH resonances of the FL protein. Further compelling evidence of interdomain flexibility has been obtained from NMR relaxation data, which are inconsistent with a rigid structure for these two full-length proteins. Actually the experimental  $R_1$  values for the residues of CAT and HPX domain in the FL protein were found sizably higher with respect to the theoretical data calculated on the respective rigid X-ray structures.<sup>3</sup> At the same time the experimental  $R_2$  values were found systematically lower than the theoretical data for the two domains in the rigid full-length proteins. The fact that the  $R_1$  values are higher and  $R_2$  lower than expected indicates intrinsic flexibility of the full length proteins and independent motion of the two domains.

The interdomain flexibility is further supported by the small NOE values found for some residues forming the linker region in both MMP1 and MMP12. The NMR analysis has been fruitfully integrated by synchrotron small-angle X-ray scattering (SAXS). Actually, the processed X-ray scattering pattern from the full length MMP1 and MMP12 can not be reproduced satisfactorily using either individual models, such as the respective experimental crystal structures, or averaging structures among open and closed conformations. For both proteins, the analysis indicates that the observed crystallographic conformations can be present also in solution, but they are flanked by a significant amount of extended conformations in equilibrium with the closed one(s).

## *Interaction between matrix metalloproteinases and extracellular matrix components :*

The current interaction study between MMPs and ECM components have provided new and relevant information regarding the structural and dynamical features of these multidomain proteins that play a role in substrate recognition and processing. The NMR data collected on FL-MMP12 and on the isolated domains have shown that elastin peptides bind the catalytic domain alone at the active site in agreement with the generally accepted idea that the latter is responsible for the elastolysis *in vivo*. However, the observed interaction *in vitro* with the HPX domain in the FL protein and its cooperative nature suggest that *in vivo* also the full length MMP12 may participate to elastolysis. Moreover, the NMR data and the docking models of the complexes between CAT-MMP12 and selected elastin fragments have showed that an extensive interaction is required for hydrolysis. In particular, our analysis indicates that the active site of MMP12 interacts with 8-11 residues belonging to the elastin cleavage site.

MMP12 and MMP1 are structurally similar. This is particularly true for CAT domains whereas the HPX domains are slightly more distinct. The structure-function relationship is a general concept in biology. Considering the structure similarity among the MMP family members, a common catalytic mechanism is expected. Therefore, the details of the catalytic mechanism revealed on CAT-MMP12 interacting with elastin can be taken as a general model

for the catalysis in all MMPs. At the same time, the preference of MMPs for the different ECM components seems mainly due to the properties of each HPX domain.

For MMP1, the HPX domain is crucial for collagenolysis.<sup>4</sup> NMR experiments performed on the FL-MMP1 protein in presence of a triple helical collagen-like model show that the HPX domain interacts with the peptide. This result was confirmed by titrating the HPX domain alone with the same peptide. Interestingly, amino-acids which experience the largest chemical shift perturbation are placed in a region poorly conserved among the MMP family.<sup>5</sup> This region may be important for the specificity of MMP1 towards triple helical collagen.

The NMR experiments have provided structural information that has been used as constraints to calculate a docking model of the HPX-THP adduct. The titration of the FL-MMP1 with THP showed that both CAT and HPX domains interact with the collagen model.

In particular, the amino-acids forming the active site of the CAT domain and those involved in the interaction in the HPX-THP experienced a sizable decrease in signal intensity. The chemical shifts changes observed for some THP resonances allowed us to identify the binding site on the collagen model. The binding mode of HPX and CAT domains on THP suggests that a reciprocal reorientation of the two domains in FL-MMP1 take place during triple helix unwinding and the following collagenolysis. To complete the model, only was missing the linker peptide which as itself constitutes a structural distance constraint for positioning the two domains of FL-MMP1. A similar approach was already described on MMP2 by removing completely the linker and using a distance constraint between extremities of the two domains.<sup>6</sup> Of course, the two MMP1 isolated domains re-associated in solution can also perform cleavage on collagen.<sup>4</sup> However, the linker peptide by keeping the two domains close, increases dramatically the enzyme efficiency by entropic contribution.

All the information collected was used to calculate a model for the complex between FL-MMP1 and THP.

The calculated structure during this research represents the first model of an MMP1-collagen-like peptide adduct and provide new information in order to clarify the mechanism of collagenolysis.

### *Perspectives*

In order to better characterise the interaction between MMP1 and the collagen peptide model, new experiments using a partially double labelled THP sample could be performed. Actually, the availability of such a sample permits a more detailed analysis of the structural features of the MMP1-THP complex allowing then to clarify the dynamical processes related to collagen unwinding.

## Reference list

1. Bertini,I., Calderone,V., Fragai,M., Jaiswal,R., Luchinat,C., Melikian,M., Mylonas,E., & Svergun,D. Evidence of reciprocal reorientation of the catalytic and hemopexin-like domains of full-length MMP-12. *J. Am. Chem. Soc.* **130**, 7011-7021 (2008).
2. Bertini,I., Fragai,M., Luchinat,C., Melikian,M., Mylonas,E., Sarti,N., & Svergun,D. Interdomain flexibility in full-length matrix metalloproteinase-1 (MMP-1). *J. Biol. Chem.* **284**, 12821-12828 (2009).
3. Iyer,S., Visse,R., Nagase,H., & Acharya,K.R. Crystal structure of an active form of human MMP-1. *Journal of Molecular Biology* **362**, 78-88 (2006).
4. Chung,L.D., Dinakarandian,D., Yoshida,N., Lauer-Fields,J.L., Fields,G.B., Visse,R., & Nagase,H. Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis. *EMBO J.* **23**, 3020-3030 (2004).
5. Massova,I., Kotra,L.P., Fridman,R., & Mobashery,S. Matrix Metalloproteinases: structures, evolution, and diversification. *FASEB J.* **12**, 1075-1095 (1998).
6. Monaco,S., Gioia,M., Rodriguez,J., Fasciglione,G.F., Di Pierro,D., Lupidi,G., Krippahl,L., Marini,S., & Coletta,M. Modulation of the proteolytic activity of matrix metalloproteinase-2 (gelatinase A) on fibrinogen. *Biochem. J.* **402**, 503-513 (2007).

## Part II :

# Study of interaction between S100P and RAGE receptor

# 1. Introduction

## *S100 family*

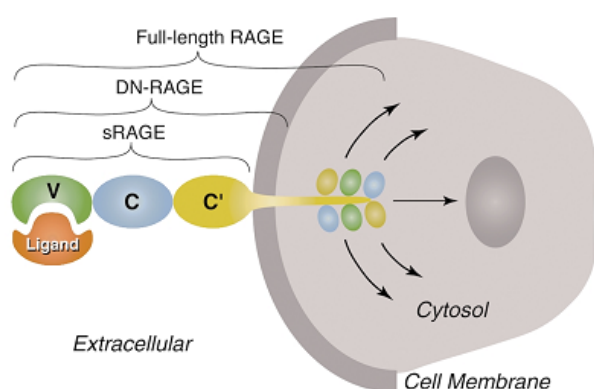
Proteins which belong to the S100 family may have an intra or an extracellular location. They are called “S100” from their 100% solubility property in a saturated solution of ammonium sulphate.<sup>1</sup> They usually consist in dimeric proteins containing two pairs of EF hand motifs. Each EF hand motif has a certain affinity to bind calcium ions. After protein activation mediated by a calcium ion binding, the S100 proteins interact with cellular receptors in order to activate different cellular functions.<sup>2</sup> In particular, two observations have been reported regarding how the S100 proteins interact with their respective targets. The first one is that the binding of a calcium ion induces conformational changes from the dimeric protein which modify in particular the relative orientation of the final alpha helix. The second observation is that S100 proteins almost always interact in a dimeric form. The consequence is that a larger contact area is obtained to interact with the partner. Sometimes, because of symmetrical properties, a S100 dimer can interact simultaneously with two partners positioned at the two opposite contact faces of the molecule. Moreover, some S100 protein can form tetramers or even octamers, increasing then their ability to bind additional molecules. This multi-binding property may have an important functional role in the activation of receptors or in signalling pathways.

Due to important differences in the physiological calcium availability inside and outside the cell, S100 proteins can play different roles depending on their sub-cellular localisation. Inside the cell, the S100 proteins can regulate protein phosphorylation by interacting with kinases often involved in signalisation pathways. S100 proteins can also inhibit some enzymes like phospholipase A2. Moreover, S100 proteins play a role in calcium homeostasis. Outside the cell, S100 proteins interact with cell surface receptors in particular with those involved in inflammation. Several S100 proteins, like S100B or S100P interacting with RAGE are responsible for activating different signalling pathways. S100 proteins expression is relatively low in mature tissues except in the tumour cell where the protein is overexpressed. In the recent years, an increased interest in the study of these proteins was developed in order to understand the activation mechanisms and to characterise the interaction between S100 and the different receptors. For example, in the case of prostate carcinoma,<sup>3</sup> S100P (originally extracted from human placenta) was highly expressed in particular in presence of an androgen.

S100P is highly expressed in neoplastic cells. The origin of this presence in tumour cells can be partially explained by an interaction between S100P and RAGE (Receptor for Advanced Glycation End products) located at the cytoplasmic membrane.<sup>4</sup> This interaction provokes via different signalling pathways such as ERK and MAP kinases an increase of cell survival and cell proliferation by stimulating S100P expression in an autocrine manner.

## Receptor for Advanced Glycation End products

RAGE which belongs to the immunoglobulin superfamily is located at the cell surface. This protein contains an extracellular region composed by a variable domain V and two constant domains C1 and C2 corresponding respectively to the variable and constant domains of the soluble immunoglobulins. In addition, RAGE contains also a transmembrane region and a cytoplasmic domain (figure 1).



**Figure 1.** Schematic depiction of RAGE-ligand interaction showing domains of the receptor. The V-type domain is critical for binding of RAGE ligands. Deletion of the cytosolic tail results in a modified form of RAGE that binds ligands, remains firmly embedded in the cell membrane, but does not transmit RAGE-mediated cellular activation. Even in the presence of endogenous full-length RAGE, expression of dominant negative RAGE blocks signalling from the receptor. sRAGE, soluble RAGE; DN-RAGE, dominant negative RAGE (extracellular domain + transmembrane-spanning domain).<sup>5</sup>

RAGE was initially identified from its ability to bind Advanced Glycation End products (AGEs) which accumulates due to diabetes or renal failure.<sup>6</sup> Indeed, AGE ligation to RAGE activates p21ras by recruiting downstream targets such as MEK and MAP kinases, followed by an activation of NF- $\kappa$ B transcription factor.

RAGE is highly expressed during the central nervous system development. In adult animal, RAGE expression level is also relatively high in lungs which suggests several functions.

To study the different functions of RAGE, a strategy has consisted in inhibiting RAGE activity by adding the extracellular soluble protein fraction also called sRAGE (figure 1) in order to compete with the natural ligand. In this way, the RAGE natural effects can be studied both *in vitro* and *in vivo* by using then different concentrations of sRAGE.<sup>5</sup>

Among the molecules binding RAGE receptor, amphoterin is a variable peptide of 10-15 amino-acid length. After amphoterin binding to the cell surface, intracellular ERK-RAGE interactions were evidenced by an enhanced ERK kinase activity constituting a starting point for the MAP kinases signalling pathway.<sup>7</sup> Else, inhibiting the interaction between RAGE and amphoterin significantly reduces this MAP kinases pathway and as a consequence tumour proliferation, invasion and expression of matrix metalloproteinases.<sup>8</sup>

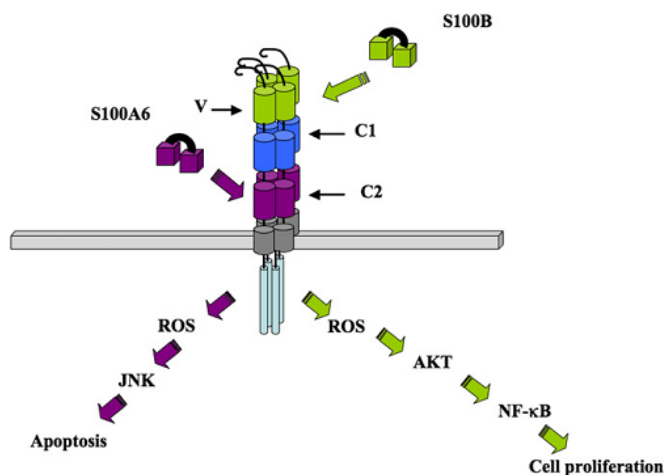


## 1.1 Pathologies related to S100P and RAGE activity

The S100 family members and RAGE receptor are involved in several pathologies such as diabetes, inflammation or cancer development. For example, increasing concentration of sRAGE has shown a significant decrease of the phenotype effects on diabetic mice indicating that RAGE is involved in diabetes.<sup>9</sup> In addition, it was reported that RAGE effects in diabetes are mediated by the NF- $\kappa$ B pathway.<sup>10</sup> In other studies, injection of sRAGE on mice accelerated the wound healing process. Else, RAGE is also involved in inflammatory process through activation by S100A12 followed by recruitment of lymphocytes and macrophages.<sup>6</sup>

RAGE also interacts with  $\alpha\beta$  amyloids.<sup>5</sup> Amyloid accumulates in particular in neurons which provokes critical troubles such as Alzheimer disease. In addition, RAGE expression is amplified by the interaction between RAGE and these  $\alpha\beta$  amyloids.

Regarding the S100 family, a gene expression profile observed on patients showed that S100P is the one genes having the highest expression fold increase in adenocarcinoma cells respect with normal conditions.<sup>11</sup> Indeed, cancer development is often connected to some changes in DNA modifying then the gene expression through an alteration of the promoter. For example, in breast cancer, lung cancer and pancreatic tumour, S100P expression is up-regulated.<sup>12</sup>



**Figure 2.** Example of RAGE signalling triggered by S100B or S100A6 in human

SH-SY5Y neuroblastomas. Micromolar concentration of S100B and S100A6 triggered distinct cellular pathways involving Akt and JNK respectively, leading to either cell survival, proliferation (S100B), or apoptosis (S100A6).<sup>13</sup>

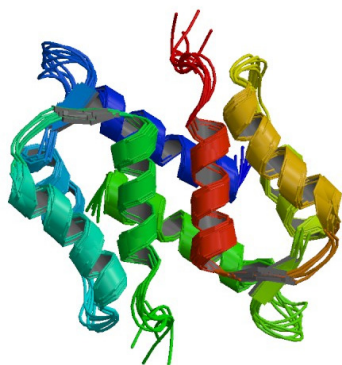
A complete interaction mechanism between S100P and RAGE was described in colon cancer. In particular S100P expression is much higher in colon cancer than in normal cell whereas the expression of RAGE is quite constant. Activation of ERK, MAP kinase pathway and further activation of the NF- $\kappa$ B transcription factor was also evidenced.<sup>14</sup>

## 1.2 Structural studies

### *S100 proteins*

S100 proteins share some similar structural pattern with calmodulin, in particular the presence of EF hands motifs which are able to bind calcium ions. Initially, calmodulin structure was solved in its apo form.<sup>15</sup> A comparison between the NMR structure of calmodulin apo form and the X-Ray structure of the  $\text{Ca}^{2+}$  complexed protein<sup>16</sup> showed some important conformational changes occur during the calcium ion binding. Later on, these conformational changes were confirmed by the NMR structure of the  $\text{Ca}^{2+}$  complexed form of calmodulin<sup>17</sup>. Evidence of the rearrangement was obtained with the contribution of residual dipolar couplings useful to study different conformations in solution.

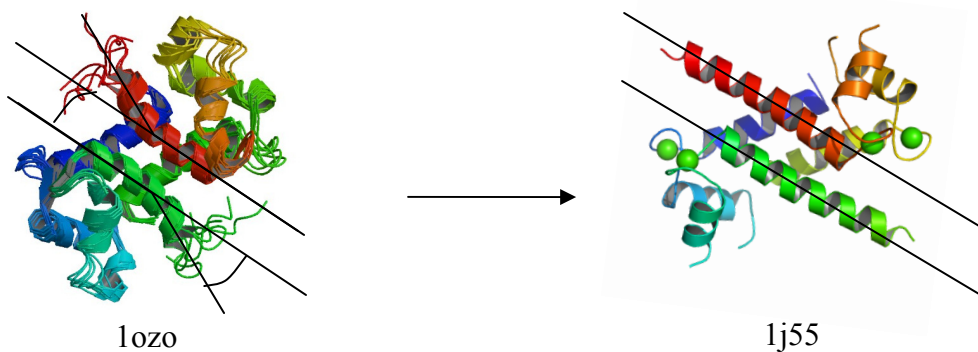
From the S100 family, NMR solution structure of the S100A6 apo form of the rabbit<sup>18</sup> was solved. Intermolecular noe constraints were used to obtain the structure assuming a dimeric form of the protein indicated by the observed line width of the signals. In a similar way, the S100B apo-protein solution structure (figure 3) from the rat was obtained.<sup>19</sup>



**Figure 3.** Solution structure of calcium-free S100B from RCSB PDB database code 1SYM.<sup>19</sup>

In the case of S100P, the use of a double hybrid approach with mutated proteins has allowed to evidence some amino-acids involved in hydrophobic contacts in the structure of the dimer.<sup>20</sup> In particular, the important role of some hydrophobic residues (F15 > F89 > I11 > I12) in the dimer formation was shown. These residues are highly conserved in the S100 family, especially the F15 which is present in all the S100 known to date.

A crystal structure was also obtained for the  $\text{Ca}^{2+}$  bound state of S100P (figure 4).<sup>21</sup> Again, the interaction between the two monomers is driven by hydrophobic contacts. Moreover, in the case of S100P, no heterodimer was observed between a S100P monomer and a different S100 family member. For other S100 members, some heterodimer like a S100A1/S100B complex has been observed.<sup>2</sup> Subsequently, the apo S100P solution structure was obtained.<sup>22</sup> The N-terminal part of the latter structure is similar to the S100P  $\text{Ca}^{2+}$  bound state structure whereas the C-terminal part shows significant changes which may suggest a role in target recognition.

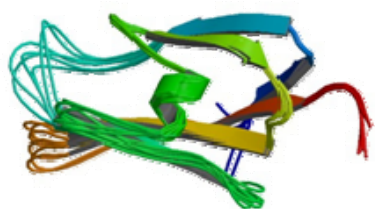


**Figure 4.** Conformational changes induced by binding of a calcium ion (green sphere) to S100P from the apo form (pdb entry 1o2o)<sup>22</sup> to the holo form (pdb entry 1j55).<sup>21</sup>

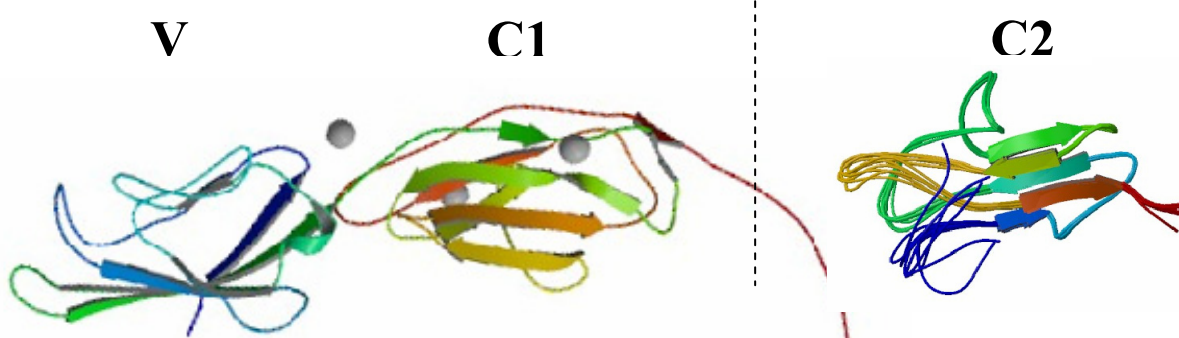
Indeed, by looking at the differences in S100 structures between the apo forms and the holo forms into the different dimers, it appears that binding of the calcium ions induces approximately a 40° angle rotation in the C-terminal  $\alpha$ -helix (figure 4). This conformational change position the EF hand in order to interact with the target.<sup>23</sup>

#### *Receptor for Advanced Glycation End products*

Structural studies were performed on RAGE receptor. Variable domain (V-domain, figure 5) was first solved by NMR (pdb entry 2e5e deposited in 2007 by Matsumoto S. et al. unpublished). Then, the second constant domain structure (C2, figure 7) was also solved by NMR (pdb entry 2ens deposited in 2008 by Wakabayashi, M. et al. unpublished). To date, the last RAGE structure available in the PDB database which consist in a V-C1 construct of RAGE was obtained by X-ray crystallography (pdb entry 3cjj deposited in 2009 by Koch M. et al. unpublished, figure 6).



**Figure 5.** Solution structure of RAGE Variable domain (pdb entry 2e5e ; Matsumoto S. et al.)



**Figure 6.** Crystal structure of RAGE construct V and C1 domains (pdb entry 3cjj ; Koch M. et al.)

**Figure 7.** Solution structure of RAGE C2 domain (pdb entry 2ens ; Wakabayashi, M. et al.)

### **1.3 Aim of the project**

The relevant pathological role of RAGE activation by the S100 family members has stimulated the research aimed to clarify the structural feature of the complexes between extracellular domains of RAGE receptor and this important class of ligands. During this research, the interaction of S100P with RAGE V-domain has been investigated by NMR. The backbone resonances of holo S100P and RAGE V-domain have been assigned. The interaction between the two proteins has been monitored by 2D HSQC and TROSY experiments. The amino-acids involved in protein-protein interaction have been identified. Then, a model for the S100P-RAGE V-domain complex has been calculated from the available structures by applying the NMR data obtained. An investigation on the dynamical properties of the complex was also carried out.

## **2. Materials and Methods**

## 2.1 NMR

### *Protein assignment experiments*

The experiments for the S100P protein assignment, mobility measurements and titration experiments were performed on protein samples at concentrations ranging between 0.4 and 1 mM. Sample conditions of S100P were usually at 75 mM of NaCl for ionic strength, 10 mM of CaCl<sub>2</sub> with a HEPES buffer at pH 7.

For RAGE V-domain, samples were concentrated in a solution containing 250 mM of NaCl, 10 mM of sodium citrate buffer up to 0.5 mM with a pH adjusted at 6.

All NMR experiments were performed and acquired on Bruker AVANCE900, AVANCE800, AVANCE700 and DRX 500 spectrometers. All instruments are equipped with a triple resonance CRYO-probes.

The assignment of the S100P-Ca<sup>2+</sup> sample was obtained by the analysis of a HSQC, <sup>24,25,26,27</sup> HNCA<sup>28</sup>, CBCA(CO)NH<sup>29</sup>, HNCACB<sup>30,31</sup> experiments performed at the 500 MHz spectrometer with a <sup>13</sup>C-<sup>15</sup>N enriched sample concentrated to 1 mM in the buffer conditions described above for S100P and at a temperature of 310K.

The backbone resonance assignment of and V-domain of RAGE receptor was obtained by the analysis of HSQC, HNCA, CBCA(CO)NH and HNCACB spectra performed at 700 MHz for HNCA and at 500 MHz spectrometer for the other 3D experiments. All these spectra were recorded at 298K.

### *NMR titration experiments*

#### 2H-13C-15N-labelled S100P with unlabelled V domain of RAGE

A triple labelled 2H-13C-15N enriched sample of S100P-Ca<sup>2+</sup> was prepared at a concentration of 0.5 mM with 75 mM of NaCl, 10 mM of CaCl<sub>2</sub> and a HEPES buffer at pH 7.

In order to study the interaction in solution by NMR between S100P and V-domain of RAGE, reference HSQC, 1H-15N trosy<sup>32</sup> TROSY-HNCA<sup>33</sup> experiments were performed at the 900 MHz at a temperature of 298K.

Then unlabelled V-domain of RAGE protein was added to the initial sample reaching a final concentration of 0.39 mM for S100P and a 0.23 mM for V-domain of RAGE. Same experiments HSQC, TROSY and HNCA were repeated on the complex at the temperature of 298K.

### 2H-13C-15N-labelled S100P with unlabelled V-C1 RAGE construct

A similar S100P-Ca<sup>2+</sup> protein sample at a concentration of 0.5 mM was prepared to study the interaction between S100P and the VC1 RAGE construct. Reference HSQC, TROSY and HNCA experiments were performed at the 900 MHz at a temperature of 310K.

Same experiments were repeated at the same temperature after addition of an unlabelled VC1 construct. Final concentrations for the complex were respectively 0.435 mM for S100P and 0.26 mM for the VC1 construct.

### 13C-15N-labelled V domain of RAGE with unlabelled S100P

To observe chemical shift changes on RAGE protein, a reciprocal titration experiment was done by using a 13C-15N double labelled sample of the RAGE V-domain protein and an unlabelled S100P-Ca<sup>2+</sup> sample. Reference HSQC and TROSY spectra of a 13C-15N labelled RAGE V-domain were acquired at the 900 MHz spectrometer at a temperature of 298K. Initial concentration of V-domain sample was 0.33 mM. Solution also contains 250 mM of NaCl and 20 mM of MES buffer at a pH of 6. Unlabelled S100P-Ca<sup>2+</sup> was progressively added to the initial sample from a 2.6 mM concentrated mother solution. Additions of several S100P amounts were done in order to reach respectively 0.5, 1 and 1.5 equivalent of S100P dimer respect with the RAGE V-domain concentration.

HSQC and TROSY spectra were performed after each addition of S100P at a temperature of 298K.

### *Relaxation experiments T1 and T2 on S100P and S100P-Vdomain complex*

#### Mobility measurements on S100P

The experiments for the determination of 15N longitudinal and transverse relaxation rates<sup>34</sup> and 1H-15N NOE<sup>35</sup> were recorded at 298 K and 500 MHz on 15N-enriched sample of S100P at a concentration of 0.5 mM. The 15N longitudinal relaxation rates (*R*<sub>1</sub>) were measured using a sequence modified to remove cross correlation effects during the relaxation delay<sup>36</sup>. Inversion recovery times used for the experiments were the following : 20, 100, 200, 300, 400, 500, 700, 900, 1100, 1300 and 1500 ms, with a recycle delay of 3.5 s. The 15N transverse relaxation rates (*R*<sub>2</sub>) were measured using a Carr-Purcell-Meiboom-Gill (CPMG) sequence<sup>37</sup> with the following delays of 8.5 , 17, 25.4, 33.9, 42.4, 50.9, 59.4, 67.8, 76.3 and 84.8 ms for the S100P protein with a refocusing delay of 450 μs.

## Mobility measurements on S100P-Vdomain complex

The same pulse sequence was used for the determination of the  $^{15}\text{N}$  longitudinal relaxation rates ( $R_1$ ) on the complex S100P-V-RAGE. Complex was obtained from a mixture of  $^{15}\text{N}$  labelled S100P with an unlabelled V-domain of RAGE with the respective concentrations of 500  $\mu\text{M}$  and 300  $\mu\text{M}$ . Solution buffer is the same as used for the corresponding titration. Inversion recovery times used for the experiments were the following : 20, 100, 200, 300, 400, 500, 700, 900, 1100, 1300, 1500, 1700, 1900, 2200 and 2500 ms, with a recycle delay of 3.5 s.

### *Effect of protein concentration in solution*

Considering, the S100P protein in solution is present as a dimer and the S100B have been shown to form tetramer or octamer. Effects of concentration on  $^{15}\text{N}$  S100P were analysed by preparing S100P samples in a similar buffer as described initially. Spectra were acquired at the 700 MHz at respectively 0.1, 0.2, 0.4, 0.7, 4 and 5 mM. Quality of spectra didn't increase above 0.4 mM indicating some line broadening effects probably due to aggregation.

Regarding S100P- RAGE V-domain complex, TROSY spectra was acquired with a diluted complex (S100P 0.4 mM : RAGE 0.35 mM) and a concentrated sample (S100P 1.2 mM : RAGE 1.05 mM). Again, the spectra quality decreases, in particular, some signals disappeared indicating some aggregation may occur.

## **2.2 Assignment of the S100P and RAGE V-domain**

Backbone assignment was performed by using CARA<sup>38</sup> programme in a similar way as described in part I (see p.34).

S100P backbone assignment of the holo form used in the present study was compared with the deposited assignment of the S100P apo form (BMRB entry 5103). Moreover, X-ray structure<sup>21</sup> of the dimeric S100P- $\text{Ca}^{2+}$  form was used to run chemical shift prediction software Sparta<sup>39</sup> and Proshift.<sup>40</sup> A table of resonances for backbone assignment is presented in the result section (see p.94).

Regarding RAGE V-domain, the solution structure has been recently resolved (pdb entry 2e5e) and then, NMR assignment is also available (BMRB entry 7364).

## **2.3 Mobility studies from relaxation data**

The same procedure as in part I (see p.35) was followed to study relaxation data. Estimates of  $R_1$  and  $R_2$  values for dimeric S100P and S100P-RAGE V-domain complex were determined by using the program HydroNMR<sup>41</sup> which simulates conditions of magnetic field and temperature from the three-dimensional structures of the proteins taken as input for the programme with respectively the X-ray structure<sup>21</sup> of the S100P- $\text{Ca}^{2+}$  and the RAGE V-domain extracted from V-C1 construct X-ray structure (pdb entry 3cjj).



## 2.4 Docking Calculation

Haddock<sup>42</sup> programme online was used to calculate the interaction model the dimeric S100P protein and V-domain of RAGE. A similar procedure was followed as in part I (see p.37). Active residues were selected from the basis on the chemical shift perturbations observed on the two partners. Considering S100P is a homodimer, selection was done first by using active residues on the two monomers and a second selection using only one monomer.

Moreover, several possible combinations of active residues were compared in order to insure the correct binding. From S100P, three zones were selected, the main zone contains the C-terminal  $\alpha$ -helices with stronger effects, the lateral part close to the linker between the two EF-hand with average effects and the opposite  $\alpha$ -helix face with smaller effects. From the RAGE V-domain structure, two zones were selected corresponding to the two faces of the proteins formed by  $\beta$ -sheets.

Crystal structures of S100P-Ca<sup>2+</sup> bound form (pdb entry 1j55) and VC1 construct of RAGE (pdb entry 3cjj) were used for docking calculation. Naccess<sup>43</sup> programme was run on the two protein structures.

Then, calculation was run assuming an automatic flexibility selection around the active residues. The different clusters obtained were analysed with molmol programme.<sup>44</sup> All the structures from each calculation run were fit by using profit programme (Martin, A.C.R. and Porter, C.T., <http://www.bioinf.org.uk/software/profit/>) on the lowest energy structure of the corresponding calculation run. The energy score function was manually defined in the following way :

$$E_{\text{score}} = 0.2 E_{\text{elec}} + E_{\text{vdw}} + E_{\text{desolv}}$$

Then, discarding the violation term of the energy score allows the comparison between different calculations with a different number of active residues selected.

## **3. Results**

### **3.1 NMR backbone assignment of S100P-Ca<sup>2+</sup> form**

Backbone assignment was obtained for the S100P-Ca<sup>2+</sup> bound form of the protein (Table 1). The protein solution structure is described as a homodimer. Indeed, the two monomers cannot be distinguished on the spectra. Despite the relatively low molecular weight of each monomer (Mw = 10 kDa), the use of a triple labelled sample allowed to increase peak resolution considering that the molecular weight for the entire protein is higher than 20 kDa. Indeed, despite a relatively regular spectral window of 7 ppm for the amide protons and 32 ppm for 15N, a large majority of the peaks found on HSQC are located in the central region of the spectra (around 1H = 8 ppm ; 15N = 120 ppm). As a result, the spectra are relatively crowded.

Nevertheless, spreading of the signals is still slightly higher than in the apo form (from assignment in bmrB entry 5103).

Then backbone assignment reached more than 90% (table 1). Missing residues are located in particular in the flexible region (G23-T25). On a total of eight non-proline unassigned residues, two serines, two threonines and one glycine are missing. These residue types are usually well identified in NMR from the C<sub>β</sub> resonance (or the absence of C<sub>β</sub> in the case of a glycine). Most of the resonances in the present assignment are comparable with those of the apo-S100P assignment (BMRB 5103) and with chemical shift prediction from SPARTA and PROSHIFT programmes. However, it was not possible to identify these missing signals. Then, all visible peaks on the spectra were assigned.

Average standard deviation of resonances between the present assignment, the apo form assignment and these two predictions were respectively 0.3 ppm for HN, 1.5 ppm for N, 0.7 ppm for CA and 0.5 ppm for CB.

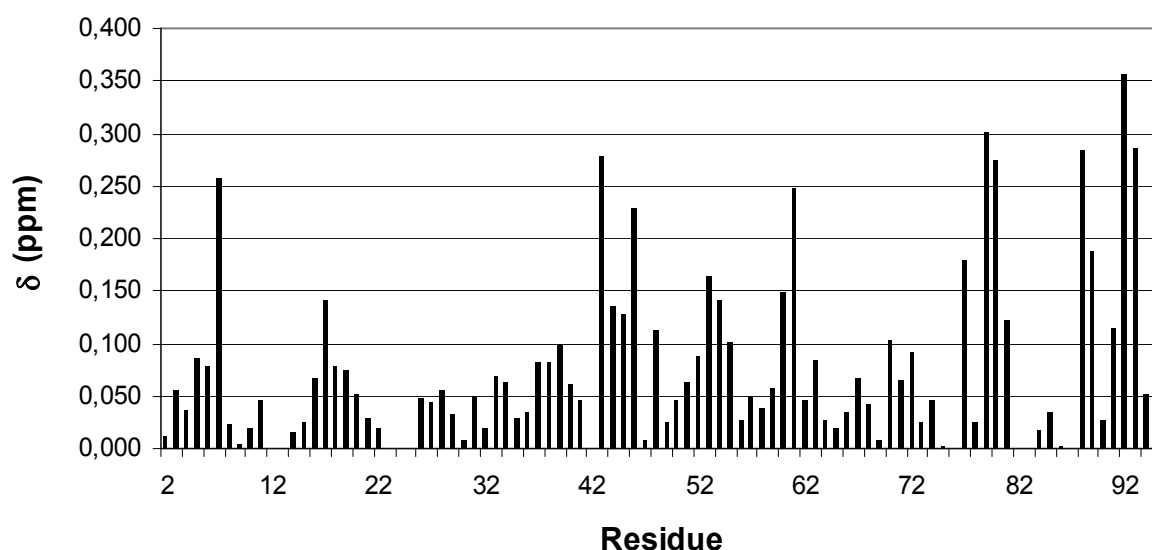
**Table 1** : NMR backbone resonances of holo S100P-Ca<sup>2+</sup> form at 298°K.

Residue	H	N	CA	CB	Residue	H	N	CA	CB
GLY	-4*				SER	47	8.00	116.55	56.00
SER	-3*		65.58		GLY	48	8.30	110.73	42.45
PHE	-2*	8.33	121.92	54.48	LYS	49	7.90	120.22	53.62 29.63
THR	-1*	8.08	116.37	58.62	ASP	50	8.00	119.27	50.90 38.10
MET	1	8.22	123.66	52.51 32.03	LYS	51	8.27	122.84	54.73 29.14
THR	2	8.95	115.11	58.21 67.04	ASP	52	8.26	119.87	51.21 37.58
GLU	3	9.53	126.15	57.63 26.22	ALA	53	7.77	123.78	52.24 16.04
LEU	4	8.78	118.74	54.77 39.40	VAL	54	8.44	118.35	62.48 27.79
GLU	5	7.89	117.66	56.52 27.42	ASP	55	7.70	121.56	54.56 37.61
THR	6	8.57	118.22	63.71 65.20	LYS	56	7.58	119.05	56.00 29.12
ALA	7	8.19	125.73	52.31 15.92	LEU	57	7.49	120.39	54.42 38.76
MET	8	8.08	115.78	56.69 29.55	LEU	58	8.27	118.32	55.70 38.12
GLY	9	8.09	105.08	43.93	LYS	59	7.36	116.62	56.02 28.83
MET	10	8.09	121.76	56.16 31.70	ASP	60	7.77	118.58	53.37 37.44
ILE	11	7.64	119.25	63.51 34.41	LEU	61	7.81	118.33	52.83 40.54
ILE	12				ASP	62	7.99	117.72	50.31 36.38
ASP	13			54.49 37.62	ALA	63	7.98	131.60	51.32 16.15
VAL	14	8.38	120.16	63.81 28.14	ASN	64	7.91	111.91	48.71 34.20
PHE	15	8.00	118.35	59.77 36.72	GLY	65	7.45	109.55	44.67
SER	16	8.50	113.22	57.60 60.29	ASP	66	8.16	120.58	50.16 37.08
ARG	17	7.89	121.91	55.56 26.38	ALA	67	10.19	122.69	50.23 12.96
TYR	18	7.01	114.49	55.95 36.90	GLN	68	7.72	113.59	50.12 29.72
SER	19	7.50	111.48	57.47 59.78	VAL	69	9.78	126.14	57.82 30.07
GLY	20	7.89	113.00	42.70	ASP	70	9.00	129.00	49.25 38.38
SER	21	7.48	117.91	57.78 60.87	PHE	71	9.19	119.84	60.08 35.70
GLU	22	9.19	120.19	51.40 30.09	SER	72	8.08	115.20	58.69 59.42
GLY	23				GLU	73	8.10	122.72	54.79 27.02
SER	24				PHE	74	8.52	120.57	57.49 37.06
THR	25			60.74 65.08	ILE	75	8.62	119.39	57.92 33.81
GLN	26	8.55	117.63	52.15 27.11	VAL	76	6.75	120.80	62.85 28.23
THR	27	7.04	104.21	55.48 70.58	PHE	77	6.78	122.95	56.19 35.59
LEU	28	9.47	124.64	49.82 40.49	VAL	78	7.78	118.60	62.82 27.86
THR	29	9.69	114.84	57.44 68.54	ALA	79	8.40	123.96	52.18 14.97
LYS	30	8.95	119.56	57.67 28.92	ALA	80	7.53	122.11	52.07 14.84
GLY	31	8.38	105.64	43.56	ILE	81	8.16	116.46	58.69 33.94
GLU	32	7.51	122.57	55.45 27.79	THR	82			
LEU	33	8.66	120.02	54.17 38.62	SER	83			58.60 59.96
LYS	34	8.00	118.61	57.22 29.64	ALA	84	7.72	123.34	51.43 15.08
VAL	35	6.98	117.69	62.94 28.35	CYS	85	7.93	117.27	58.26 23.79
LEU	36	7.64	120.98	56.26 38.61	HIS	86	7.91	120.85	56.61
MET	37	8.36	117.00	54.01 27.28	LYS	87			54.99 27.36
GLU	38	7.92	116.89	55.82 26.70	TYR	88	7.77	119.16	56.49 35.01
LYS	39	8.01	114.74	54.00 30.33	PHE	89	7.71	118.86	55.67 35.73
GLU	40	8.75	114.77	53.50 27.78	GLU	90	7.81	121.19	54.45 26.85
LEU	41	7.71	120.06	49.28 40.08	LYS	91	7.81	120.82	53.56 29.38
PRO	42			61.34 28.03	ALA	92	7.92	123.47	49.69 15.89
GLY	43	8.68	109.95	42.34	GLY	93	7.98	107.39	42.13
PHE	44	8.03	122.33	57.70 36.76	LEU	94	7.80	121.84	52.06 39.23
LEU	45	8.37	119.03	51.55 38.94	LYS	95	7.66	126.70	54.31 30.51
GLN	46	8.12	121.34	52.50 25.32					
					percentage assigned	H 91%	N 91%	CA 96%	CB 94%

\*Correspond to the rest of the tag after cleavage

### 3.2 Interaction between S100P protein and V-domain of RAGE receptor

Titration experiments were performed using a 2H-13C-15N-labelled S100P sample (0.5 mM) in presence of an unlabelled RAGE V-domain sample. Several chemical shifts changes were observed on the spectra indicating that the two proteins interact (figure 8).

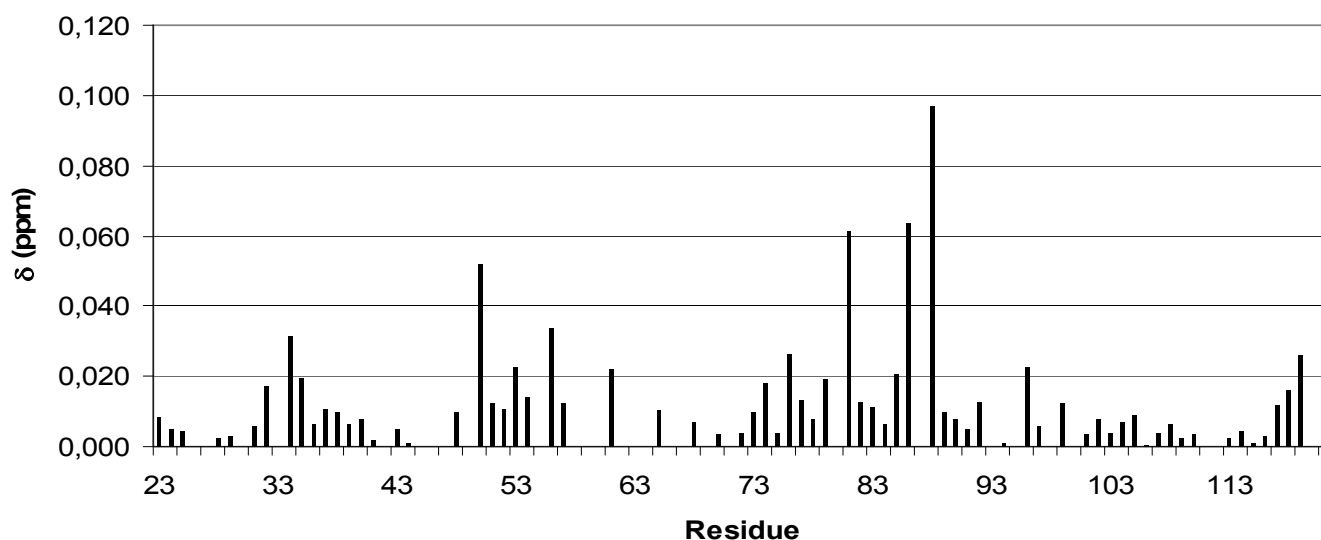


**Figure 8.** Chemical shift changes on 2H-15N-13C labelled S100P sample by titration with an unlabelled RAGE V-domain sample. Chemical shift differences represented correspond to the sum of absolute value of 1H chemical shift difference and 0.2 \* absolute value of N chemical shift difference between the reference spectra and in the presence of RAGE V-domain. Experimental conditions : [S100P]<sub>monomer</sub> = 0.39 mM ; [V-RAGE] = 0.23 mM ; Magnetic Field = 21T (900 MHz) ; T = 298K ; 2D Trosy experiment.

From the titration of 2H-13C-15N labelled S100P by the unlabelled RAGE V-domain, several changes can be observed. Some signals became broader, other simply shifted or disappeared. The stronger effects correspond to the amino-acids located in the  $\alpha$ -helix at the C-terminal on EF-Hand 2 (residues 75-95). Many changes are observed also in the central part of the sequence in particular close to the linker peptide between the two EF-hands (residues 43-55).

Additionally, the titration between 2H-13C-15N S100P and the unlabelled V-C1 construct of RAGE showed the same region experienced chemical shift changes respect with the previous titration performed on RAGE V-domain. Consequently, this result indicates that the RAGE constant domain C1 is probably not involved in the interaction with S100P. This observation is consistent with the situation observed in the interaction between an antigen and an immunoglobulin in which constant domains are not involved.

From the previous result, the RAGE V-domain was selected to study the interaction between RAGE and S100P. To confirm this interaction and to collect chemical shift changes on the RAGE V-domain, a titration was performed by using a  $^{15}\text{N}$ -labelled RAGE V-domain in presence of an unlabelled S100P sample (figure 9).



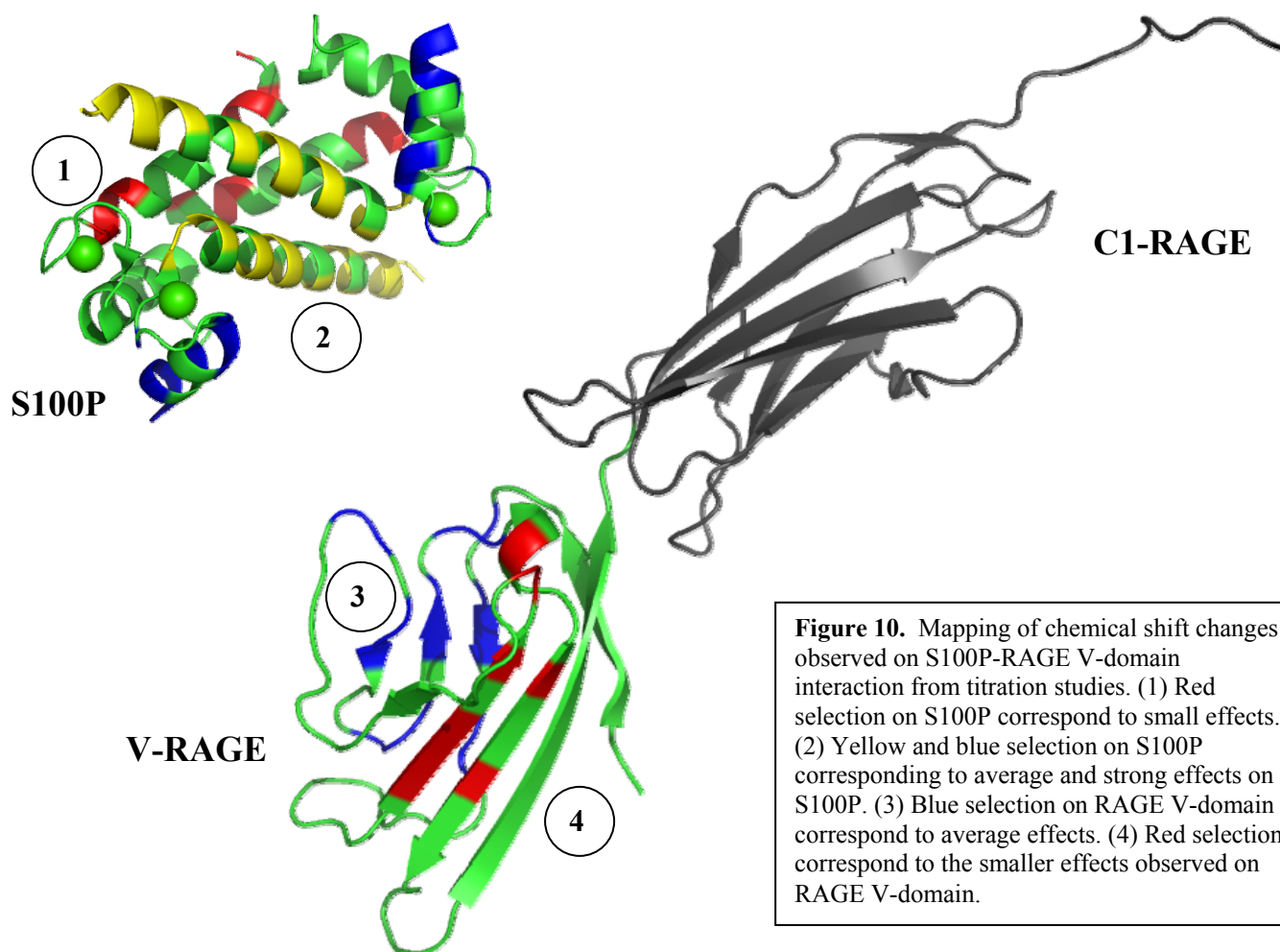
**Figure 9.** Chemical shift changes on  $^{15}\text{N}$  labelled RAGE V-domain sample by titration with an unlabelled S100P sample. Chemical shift differences represented correspond to the sum of absolute value of  $^1\text{H}$  chemical shift difference and  $0.2 \times$  absolute value of  $^{15}\text{N}$  chemical shift difference between the reference spectra and in the presence of S100P. Experimental conditions :  $[\text{V-RAGE}] = 0.33 \text{ mM}$  ;  $[\text{S100P}]_{\text{monomer}} = 0.5 \text{ mM}$  ; Magnetic Field = 21T (900 MHz) ; T = 310K ; 2D Trosy experiment.

Chemical shift changes observed were more modest comparing with the shifts observed previously between the triple labelled S100P and unlabelled RAGE V-domain. In the other hand, signal to noise ratio was better in this case and ambiguity was also lower considering RAGE V-domain is a monomer in solution. The most intense effects were observed for the amino-acids located in the central part of the protein corresponding to a  $\beta$ -sheet surface (residues 81, 86 and 88). Few effects are visible on the opposite face of the receptor probably due to a perturbation in the  $\beta$ -bonding provoked by the interaction with the partner.

These titration experiments allowed a better understanding of the regions involved in the interaction from the two partners.

### 3.3 Docking calculation on a S100P-RAGE V-domain complex

From the results obtained from titration experiments, amino-acids involved in the interaction were selected in order to drive the docking with Haddock programme. Chemical shift mapping of residues from changes observed on S100P and RAGE V-domain is reported in Figure 10.



Docking Combinations Energy score (kCal.mol <sup>-1</sup> )		V-RAGE selection		
		(3) (blue)	(4) (red)	(3) + (4)
S100P selection	(1) (red)	-133.4	-134.1	-87.9
	(2) (yellow – blue)	<b>-240.4</b>	-154.5	-221.1
	(1) + (2)	-226.0	-157.1	-162.2
	Monomer (2)	-140.4	-170.3	-133.9

**Table 2.** Energy score obtained on docking calculations using different selections from both partners.

Results obtained from docking calculations are reported in table 2. For all these different selections, an “energy score” was defined from the relation :

$$E_{\text{score}} = 0.2 E_{\text{elec}} + E_{\text{vdw}} + E_{\text{desolv}}$$

According to this energy score, the best result was obtained from the selection constituted by the residues coloured in yellow and blue on the figure (EF-Hand 2 from S100P structure) from the S100P dimer and by the residues coloured in blue from the RAGE V-domain. The score reported in the table 2 refers each time to the best cluster from docking calculation. Interestingly, from the RAGE V-domain all the docking models based on the red selection (column (4) red from the table 2) have a significantly higher energy score respect with the best model obtained (bold in the table 2). The same observation can be done with the red selection from the S100P dimer (row (1) red from the table 2). Eventually, docking control calculations performed by using a single monomer showed also a significantly worse energy score respect to the best calculation according to energy score (bold). The result from these calculations is somehow expected considering the selection from the most significant chemical shift changes observed corresponds to the best model obtained. Nevertheless, all combinations needed to be valuated to ensure finding the best matching possible. The table below contains the detail of the different selections.

---

S100P\*–Yellow (EF-Hand 2 motif, helix 4)

D70, F71, S72, V76, F77, A79, A80, I81, A84, H86, Y88, F89, E90, K91, A92, G93, L94

S100P\*–Blue (EF-Hand 2 motif, helix 3)

D52, A53, V54, D55, L57, D60, L61, A63, A67

S100P\*–Red (N-terminal region, helix 1)

M1, E3, E5, T6, A7, S16, R17, Y18, S19

V-RAGE-Blue

E32, L34, V35, K37, C38, D73, S74, A76, R77, N81, G82, S83, F85, L86, A88, V89

V-RAGE-Red

E50, W51, K52, N54, G56, R57, Q92, I96, C99

**Table 3.** Residues selected for docking calculations; symbol \* indicates a selection for both monomers except contrary indication.

---

Another element to characterise the docking is the contact area between the two partners (table 4). Usually, in interactions, the contact area is correlated with the quality of the energy function. Hence, the following table reports the contact area for these different models.

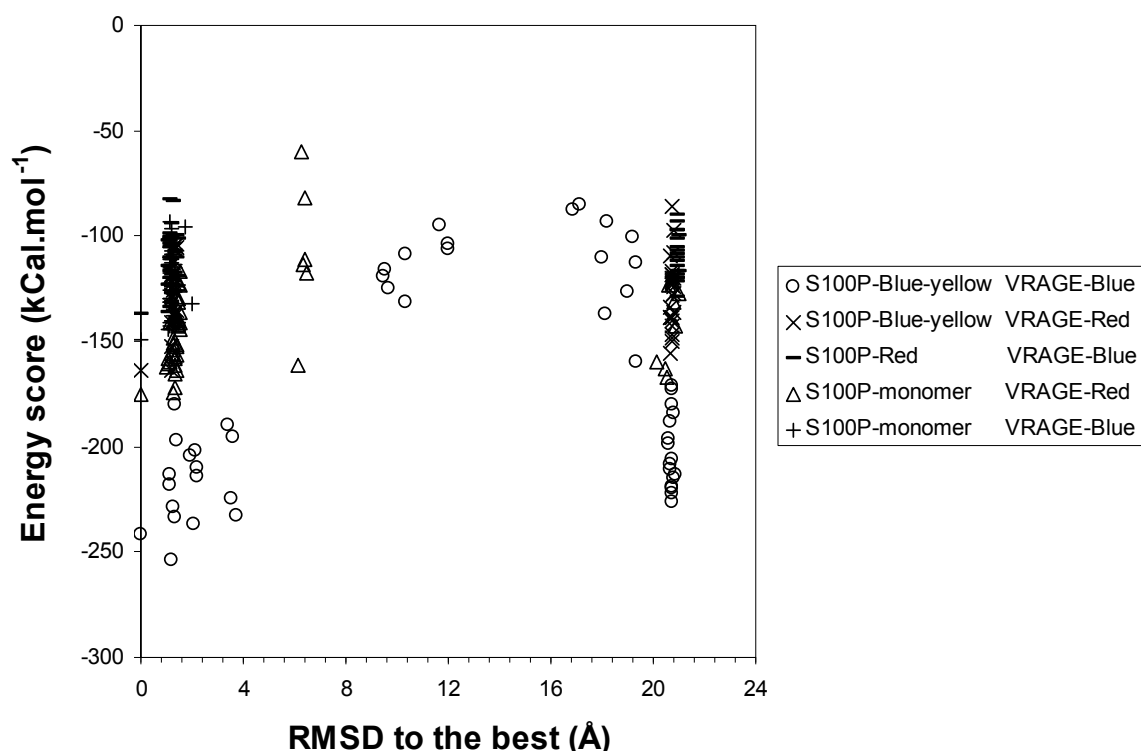
Docking Combinations Contact area (Å <sup>2</sup> )		V-RAGE selection		
		(3) (blue)	(4) (red)	(3) + (4)
S100P selection	(1) (red)	1841	1566	1742
	(2) (yellow – blue)	<b>3311</b>	2983	3155
	(1) + (2)	3212	3034	<b>3320</b>
	Monomer (2)	2292	2351	2311

**Table 4.** Contact area obtained on docking calculations using different selections from both partners.



Comparison between the contact areas of these different models showed significant differences respect with the best models from the energy score. In particular, a strong dependency can be noticed from the residue selection of S100P. It appears clearly, the interaction area between the two partners is significantly lower with the red selection of S100P (EF-Hand 1 motif) respect with the yellow-blue selection (EF-Hand 2 motif). As expected, the surface area is significantly lower on the selection from the monomer.

Combining the information from the energy score and the surface area led us to consider the interaction model obtained with S100P yellow-blue selection (EF-Hand 2 motif) and V-RAGE blue selection as a model of reference for comparison.



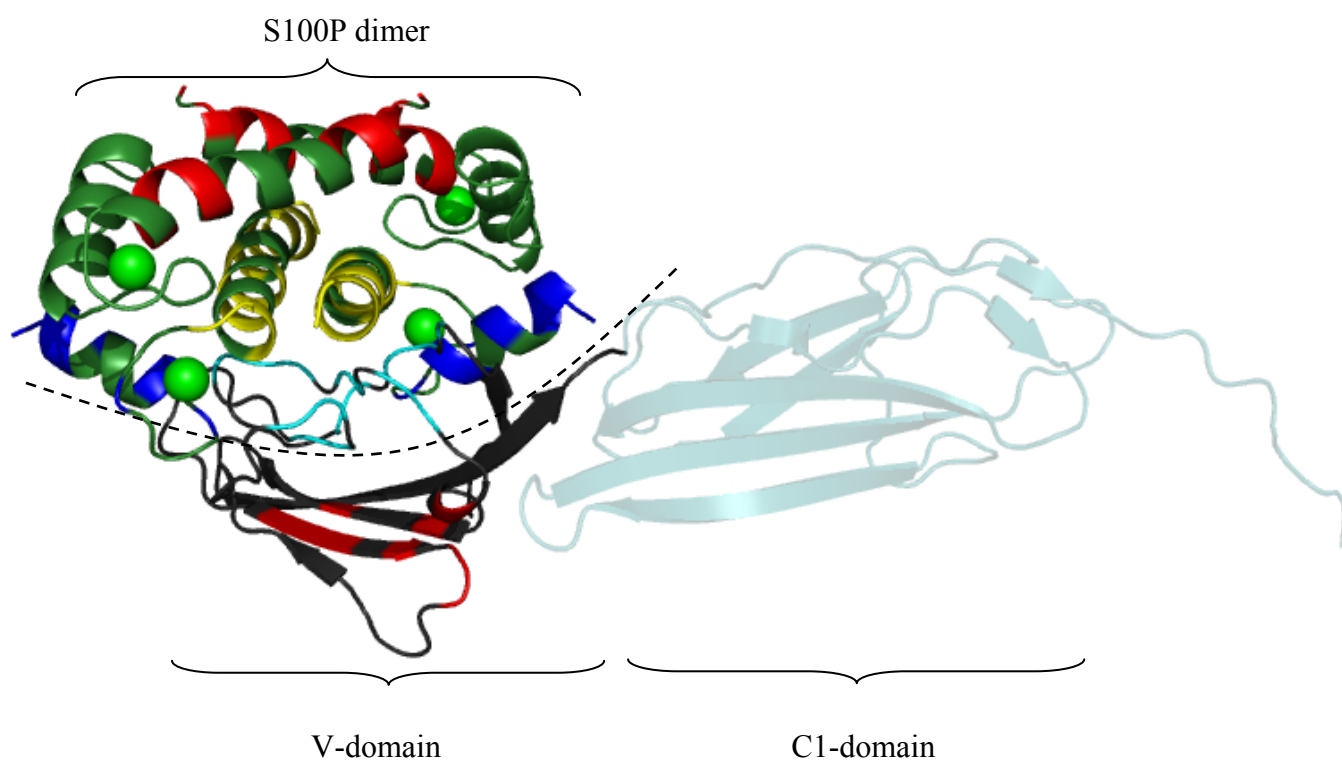
**Figure 11.** Docking calculation results from the different run. Each model was compared to the best from the corresponding calculations. The different clusters can be represented on this plot.

On the plot reported in Figure 11, it appears the calculation giving the lowest energy score (circles on Figure 11) is divided into four clusters. The best two clusters of the S100P – Blue Yellow (EF-Hand 2 motif) selection with V-RAGE Blue are comprised in an energy range between  $-180 \text{ kCal.mol}^{-1}$  and  $-250 \text{ kCal.mol}^{-1}$ . In the other hand, one cluster has an average RMSD to the best value of  $1.2 \text{ \AA}$  while the other one has an average value of  $20.7 \text{ \AA}$ . Curiously, this “away” cluster from the best structure has similar energy statistics values respect to the closer one.

By displaying these two clusters on molmol, it appears they are actually identical but the positions of the two homodimers of S100P from calculations are inverted. In fact, RMSD fitting doesn't recognise the similarity of the two monomers. The presence of these “twin cluster” is observed in all the different calculation with comparable populations.

All the other calculations generated models with an energy score comprised in a range between  $-80 \text{ kCal.mol}^{-1}$  and  $-180 \text{ kCal.mol}^{-1}$  which are all higher than the best two clusters described above.

An interaction model between S100P and RAGE V-domain from the best is a represented here.



**Figure 12.** Interaction model from docking calculation between a S100P dimer forming a complex with RAGE V-domain.

From this model obtained (figure 12), the interaction involved mainly the two  $\alpha$ -helices from EF-Hand 2 on S100P and a relatively disordered region of RAGE V-domain. Interestingly, these helices of S100P correspond to the ones experiencing some conformational changes during the calcium ion binding.

### 3.4 Effects of complex formation on dynamics

In order to demonstrate the effects of interaction on S100P protein mobility, relaxation measurements were done on a S100P protein sample concentrated at 0.5 mM (figures 13a and 13b).

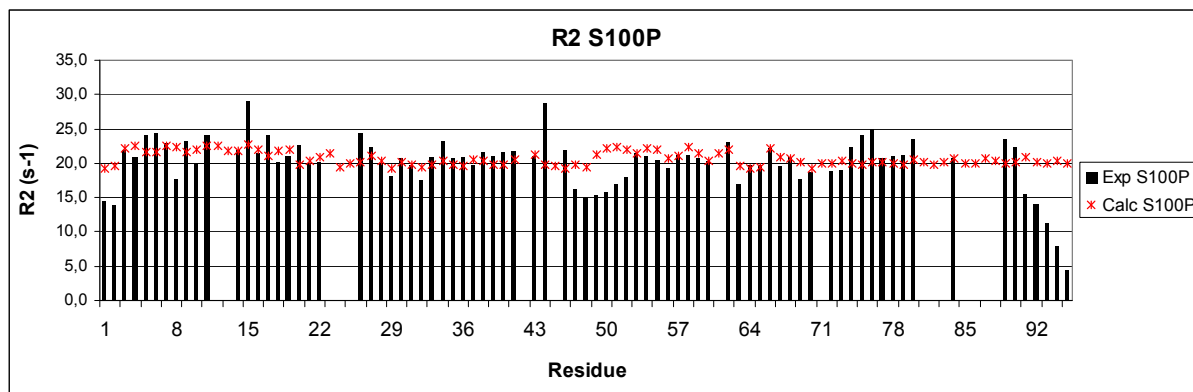


Figure 13a. Relaxation rate R2 measurements of isolated S100P protein on 500 MHz spectrometer at 298K.

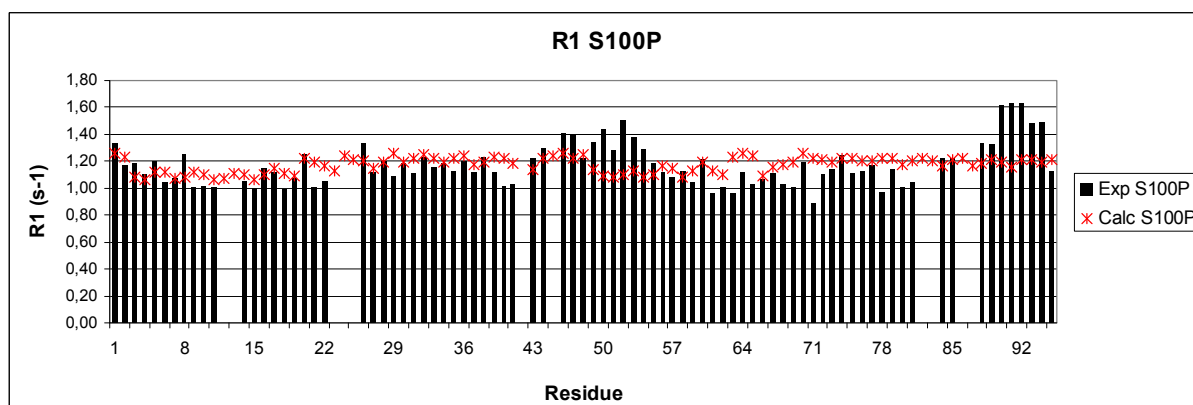


Figure 13b. Relaxation rate R1 measurements of isolated S100P protein 500 MHz spectrometer at 298K.

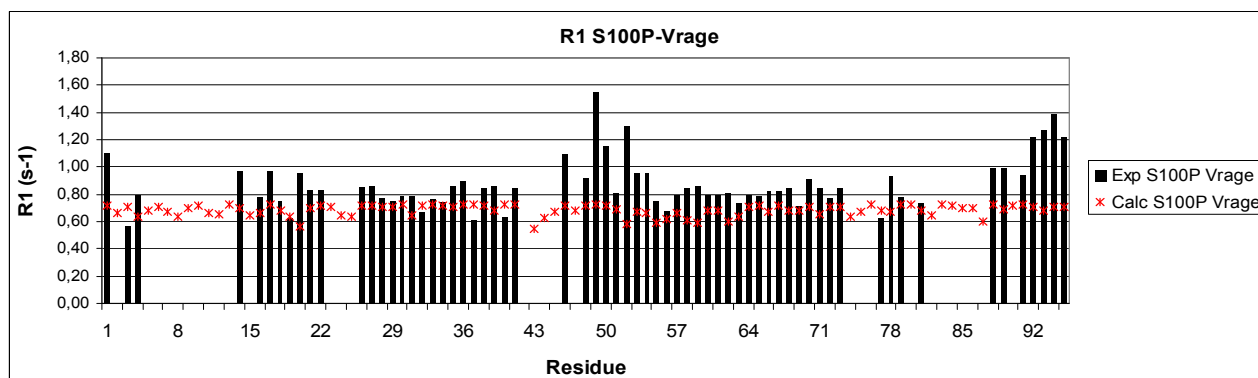


Figure 13c. Relaxation rate R1 measurements of S100P-RAGE V-domain complex on 500 MHz spectrometer at 298K.

From X-ray structure 1J55, the atomic coordinates for amino-acids contained in the loop between Q46 – K51 are missing. Interestingly, the authors reported possible mobility for the linker between the two EF-hands.

Indeed, from the R2 and R1 plots (figures 13a and 13b), this loop presents an enhanced mobility respect with the rest of the protein. Calculated values were obtained using a model derived from the dimeric X-ray structure where the missing loop was rebuilt by modeller. Relaxation values were simulated by HydroNMR assuming the protein as a rigid body. Here, experimental R2 is lower than calculated for the Q46 – K51 loop and experimental R1 is higher for the same region. This indicates that this loop displays a local flexibility.

Relaxation studies showed also the six C-terminal residues (E90-K95) present a significant mobility. Again R2/R1 ratio is lower respect with the rest of the protein. For these residues located close to the C-terminal extremity such mobility could be expected but in the other hand they are located in a  $\alpha$ -helix secondary structure according to the X-Ray structure. For all the other residues, experimental and calculated data are matching confirming the protein is present as a dimer in solution. These data obtained on isolated S100P can be used as a reference to analyse data obtained the complex. It also validates the estimation of relaxation rates by HydroNMR.

Relaxation measurements were performed on S100P-RAGE Vdomain complex (figure 13c). Results obtained on T1 measurements showed, as expected, a general decrease of R1 values for S100P residues indicating the formation of a bigger molecule. Interestingly, the mobility observed in the middle loop and at the C-terminal helix is retained. Most probably, these regions are not directly involved in close contact with V-domain but might participate in some conformational change. Calculated data were obtained using HydroNMR programme from a Haddock model of the S100P dimer-V-RAGE complex. Again, for the rest of the protein, experimental and calculated R1 data are matching confirming the complex in solution corresponds to a S100P homodimer with RAGE V-domain at the concentration used (0.5 mM for S100P monomer and 0.35 mM for V-domain). Although the structure of the interaction between S100P and RAGE V-domain has not been solved, mobility studies indicated that relaxation data are consistent with our model.

## **4. Conclusion**

### *Interaction studies of a S100P-RAGE V-domain complex.*

Experimental data show that S100P is a ligand of RAGE. The S100P protein structures in its apo form in solution by NMR (1OZO)<sup>22</sup> and in its holo form by X-ray crystallography (1J55),<sup>21</sup> are dimers. Binding of the calcium ion induces some conformational change which modifies the interface between the two monomers. As reported,<sup>4</sup> only the calcium bound form of S100P is able to interact with V-domain. Interestingly, the most significant chemical shift changes observed on S100P in presence of RAGE V-domain correspond to the S100P C-terminal residues where the structure is modified during the binding of calcium ion.

Docking calculations were performed in order to find the best adduct using the chemical shift changes obtained from the titration experiments. The best docking model has been obtained considering energy values corresponding to the protein regions from the two partners in which the chemical shift perturbations were the most significant. Moreover, the same docking model had also the most extended contact area between the two partners.

A series of relaxation time measurements was done on the holo-S100P alone and in complex with RAGE V-domain. Experiments on isolated holo-S100P indicated some flexibility in the linker between the two EF-hands of the protein and at the C-terminal region. These R1 and R2 measurements confirmed that S100P at 0.5 mM in solution is present as a homodimer. Experimental relaxation values were in agreement with the simulated values from HydroNMR programme. Indeed, most of the S100 family members are present in solution<sup>2</sup> in a dimeric form. However, S100B which also interacts with RAGE can be found as tetramer or octamer in solution.<sup>45</sup> Possibility of a tetramer formation was also simulated by the programme but results obtained were not consistent with the experimental relaxation values at this concentration.

Comparison between relaxation data measured on S100P alone and in the presence of V-domain confirmed the formation of the complex constituted by a homodimer S100P with a monomeric RAGE V-domain. Moreover, the experimental relaxation data were consistent with calculated data from the model using HydroNMR. This agreement constitutes an important step to understand the interaction properties between S100P and RAGE.

### *Perspectives*

Further studies should be carried out in order to describe the structure of the S100P-RAGE V-domain complex in more detail. For example, experiments to obtain some distance constraints from intermolecular NOE can be performed. Another strategy would consist in mutating one of the calcium binding sites on S100P to substitute one of the calcium ions by a paramagnetic metal ion. In this way, a titration performed between a <sup>15</sup>N-labelled RAGE V-domain and a paramagnetic unlabelled S100P would allow identifying paramagnetic effect on RAGE V-domain. Such effects can provide new distance constraints between S100P and V-domain useful in docking calculations.

## 4.1 Reference list

1. Moore,B.W. A soluble protein characteristic of the nervous system. *Biochem. Biophys. Res. Commun.* **19**, 739-744 (1965).
2. Donato,R. S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles. *Int. J. Biochem. Cell Biol.* **33**, 637-668 (2001).
3. Averboukh,L., Liang,P., Kantoff,P.W., & Pardee,A.B. Regulation of S100P expression by androgen. *Prostate* **29**, 350-355 (1996).
4. Arumugam,T., Simeone,D.M., Schmidt,A.M., & Logsdon,C.D. S100P stimulates cell proliferation and survival via receptor for activated glycation end products (RAGE). *J. Biol. Chem.* **279**, 5059-5065 (2004).
5. Schmidt,A.M., Yan,S.D., Yan,S.F., & Stern,D.M. The multiligand receptor RAGE as a progression factor amplifying immune and inflammatory responses. *J. Clin. Invest* **108**, 949-955 (2001).
6. Hofmann,M.A., Drury,S., Fu,C., Qu,W., Taguchi,A., Lu,Y., Avila,C., Kambham,N., Bierhaus,A., Nawroth,P., Neurath,M.F., Slattery,T., Beach,D., McClary,J., Nagashima,M., Morser,J., Stern,D., & Schmidt,A.M. RAGE mediates a novel proinflammatory axis: a central cell surface receptor for S100/calgranulin polypeptides. *Cell* **97**, 889-901 (1999).
7. Ishihara,K., Tsutsumi,K., Kawane,S., Nakajima,M., & Kasaoka,T. The receptor for advanced glycation end-products (RAGE) directly binds to ERK by a D-domain-like docking site. *FEBS Lett.* **550**, 107-113 (2003).
8. Taguchi,A., Blood,D.C., del Toro,G., Canet,A., Lee,D.C., Qu,W., Tanji,N., Lu,Y., Lalla,E., Fu,C., Hofmann,M.A., Kislinger,T., Ingram,M., Lu,A., Tanaka,H., Hori,O., Ogawa,S., Stern,D.M., & Schmidt,A.M. Blockade of RAGE-amphoterin signalling suppresses tumour growth and metastases. *Nature* **405**, 354-360 (2000).
9. Bucciarelli,L.G., Wendt,T., Rong,L., Lalla,E., Hofmann,M.A., Goova,M.T., Taguchi,A., Yan,S.F., Yan,S.D., Stern,D.M., & Schmidt,A.M. RAGE is a multiligand receptor of the immunoglobulin superfamily: implications for homeostasis and chronic disease. *Cell Mol. Life Sci.* **59**, 1117-1128 (2002).
10. Bierhaus,A., Schiekofe,S., Schwaninger,M., Andrassy,M., Humpert,P.M., Chen,J., Hong,M., Luther,T., Henle,T., Kloting,I., Morcos,M., Hofmann,M., Tritschler,H., Weigle,B., Kasper,M., Smith,M., Perry,G., Schmidt,A.M., Stern,D.M., Haring,H.U., Schleicher,E., & Nawroth,P.P. Diabetes-associated sustained activation of the transcription factor nuclear factor-kappaB. *Diabetes* **50**, 2792-2808 (2001).
11. Beer,D.G., Kardia,S.L., Huang,C.C., Giordano,T.J., Levin,A.M., Misek,D.E., Lin,L., Chen,G., Gharib,T.G., Thomas,D.G., Lizyngess,M.L., Kuick,R., Hayasaka,S., Taylor,J.M., Iannettoni,M.D., Orringer,M.B., & Hanash,S. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat. Med.* **8**, 816-824 (2002).
12. Marenholz,I., Heizmann,C.W., & Fritz,G. S100 proteins in mouse and man: from evolution to function and pathology (including an update of the nomenclature). *Biochem. Biophys. Res. Commun.* **322**, 1111-1122 (2004).
13. Leclerc, E., Fritz, G., Vetter, S. W., and Heizmann, C. W. Binding of S100 proteins to RAGE: An update. *Biochim.Biophys.Acta* . 2009.  
Ref Type: In Press
14. Fuentes,M.K., Nigavekar,S.S., Arumugam,T., Logsdon,C.D., Schmidt,A.M., Park,J.C., & Huang,E.H. RAGE activation by S100P in colon cancer stimulates growth, migration, and cell signaling pathways. *Dis. Colon Rectum* **50**, 1230-1240 (2007).

15. Kuboniwa,H., Tjandra,N., Grzesiek,S., Ren,H., Klee,C.B., & Bax,A. Solution structure of calcium-free calmodulin. *Nature Struct. Biol.* **2**, 768-776 (1995).
16. Zhang,M., Tanaka,T., & Ikura,M. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nat Struct Biol* **2**, 758-767 (1995).
17. Chou,J.J., Li,S., Klee,C.B., & Bax,A. Solution structure of Ca<sup>2+</sup> calmodulin reveals flexible hand-like properties of its domains. *Nature Struct. Biol.* **8**, 990-997 (2001).
18. Potts,B.C., Smith,J., Akke,M., Macke,T.J., Okazaki,K., Hidaka,H., Case,D.A., & Chazin,W.J. The structure of calcyclin reveals a novel homodimeric fold for S100 Ca(2+)-binding proteins. *Nat. Struct. Biol.* **2**, 790-796 (1995).
19. Drohat,A.C., Amburgey,J.C., Abildgaard,F., Starich,M.R., Baldisseri,D., & Weber,D.J. Solution structure of rat apo-S100B(beta beta) as determined by NMR spectroscopy. *Biochemistry* **35**, 11577-11588 (1996).
20. Koltzsch,M. & Gerke,V. Identification of hydrophobic amino acid residues involved in the formation of S100P homodimers in vivo. *Biochemistry* **39**, 9533-9539 (2000).
21. Zhang,H., Wang,G., Ding,Y., Wang,Z., Barraclough,R., Rudland,P.S., Fernig,D.G., & Rao,Z. The crystal structure at 2Å resolution of the Ca<sup>2+</sup> -binding protein S100P. *J. Mol. Biol.* **325**, 785-794 (2003).
22. Lee,Y.C., Volk,D.E., Thivyanathan,V., Kleerekoper,Q., Gribenko,A.V., Zhang,S., Gorenstein,D.G., Makhatazde,G.I., & Luxon,B.A. NMR structure of the Apo-S100P protein. *J. Biomol. NMR* **29**, 399-402 (2004).
23. Santamaria-Kisiel,L., Rintala-Dempsey,A.C., & Shaw,G.S. Calcium-dependent and -independent interactions of the S100 protein family. *Biochemical Journal* **396**, 201-214 (2006).
24. Bodenhausen,G. & Ruben,D.J. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69**, 185-188 (1980).
25. Piotto,M., Saudek,V., & Sklenar,V. Gradient-tailored excitation for single quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **2**, 661-666 (1992).
26. Sklenar,V., Piotto,M., Leppik,R., & Saudek,V. Gradient-tailored water suppression for <sup>1</sup>H-<sup>15</sup>N HSQC experiments optimized to retain full sensitivity. *J. Magn. Reson. Ser. A* **102**, 241-245 (1993).
27. Mori,S., Abeygunawardana,C., Johnson,M.O., & van Zijl,P.C. Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J. Magn Reson. B* **108**, 94-98 (1995).
28. Grzesiek,S. & Bax,A. Improved 3D Triple-Resonance NMR Techniques Applied to a 31 KDa Protein. *J. Magn. Reson.* **96**, 432-440 (1992).
29. Farrow,N.A., Muhandiram,R., Singer,A.U., Pascal,S.M., Kay,C.M., Gish,G., Shoelson,S.E., Pawson,T., Forman-Kay,J.D., & Kay,L.E. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by <sup>15</sup>N NMR relaxation. *Biochemistry* **33**, 5984-6003 (1994).
30. Grzesiek,S. & Bax,A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.* **99**, 201-207 (1992).
31. Muhandiram,D.R. & Kay,L.E. Gradient-enhanced triple resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. Ser. B* **103**, 203-216 (1994).
32. Czisch,M. & Boelens,R. Sensitivity enhancement in the TROSY experiment. *J. Magn Reson.* **134**, 158-160 (1998).



33. Salzmann, M., Pervushin, K., Wider, G., Senn, H., & Wüthrich, K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. Natl. Acad. Sci. USA* **95**, 13585-13590 (1998).
34. Palmer, A.G., III, Rance, M., & Wright, P.E. Intramolecular motions of a zinc finger DNA-binding domain Xfin characterized by proton-detected natural abundance <sup>13</sup>C heteronuclear NMR spectroscopy. *J. Am. Chem. Soc.* **113**, 4371-4380 (1991).
35. Zhu, G., Xia, Y., Nicholson, L.K., & Sze, K.H. Protein dynamics measurements by TROSY-based NMR experiments. *J Magn Reson* **143**, 423-426 (2000).
36. Kay, L.E., Nicholson, L.K., Delaglio, F., Bax, A., & Torchia, D.A. Pulse sequences for removal of the effects of cross correlation between dipolar and chemical-shift anisotropy relaxation mechanisms on the measurement of heteronuclear T<sub>1</sub> and T<sub>2</sub> values in proteins. *J. Magn. Reson.* **97**, 359-375 (1992).
37. Peng, J.W. & Wagner, G. Investigation of protein motions via relaxation measurements. *Methods Enzymol.* **239**, 563-596 (1994).
38. Keller, R. L. J. *The Computer Aided Resonance Assignment Tutorial*. [1.3]. 2004. CANTINA Verlag. Ref Type: Computer Program
39. Shen, Y. & Bax, A. Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J. Biomol. NMR* **38**, 289-302 (2007).
40. Meiler, J. PROSHIFT: protein chemical shift prediction using artificial neural networks. *J. Biomol. NMR* **26**, 25-37 (2003).
41. de la Torre, J.G., Huertas, M.L., & Carrasco, B. HYDRONMR: Prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* **147**, 138-146 (2000).
42. Dominguez, C., Boelens, R., & Bonvin, A.M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731-1737 (2003).
43. Lee, B. & Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **55**, 379-400 (1971).
44. Koradi, R., Billeter, M., & Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structure. *J. Mol. Graphics* **14**, 51-55 (1996).
45. Ostendorp, T., Leclerc, E., Galichet, A., Koch, M., Demling, N., Weigle, B., Heizmann, C.W., Kroneck, P.M., & Fritz, G. Structural and functional insights into RAGE activation by multimeric S100B. *EMBO J.* **26**, 3868-3878 (2007).

## ***Acknowledgments***

*I am sincerely grateful to:*

*Professor Ivano Bertini who gave me the opportunity to carry out my PhD in Structural Biology at CERM,*

*Professor Claudio Luchinat, for the interesting scientific discussions we had together and the support a gave me to attend conferences,*

*Dr. Marco Fragai for his guidance throughout this nice project, for all the constant support he gave me in all these challenging projects.*

*Dr. Massimiliano Peana who introduced me into the NMR world.*

*Drssa. Chiara Venturi, Linda Cerofolini, Rahul Jaiswal, Niko Sarti, Mirco Toccafondi, Valenti Borsi, Antonella Nesi and Joao Teixeira for all the nice experiments and all the work we did together.*

*I also want to bring special thanks to:*

*My parents for all their love and constant support in particular in the difficult moments, to all my family and my friends from France, in particular my best friend Fabien.*

*My dear colleagues, in particular Malini, Ina, Shailesh... We spent really nice moments together.*

*“I ragazzi” of the international CERM football team who helped me to get some relaxation after work.*

*All the boys and girls who passed by Val di Rose, the house where I lived during my PhD thesis. : Anna, Luca, Vlad, Ala, Veronika, Rossella, Fiore, Jasper, Maciek, Marcin, Andris, Yanis, Michele, Vaishali, Deepa, Mathieu, Benoit, Isabelle, Mao and Fuyuan. Special thanks to Marco. We had a lot of fun together in particular watching football games and filmoni. Special thanks also to mia cara Mirela for all the nice moments we have spent together and for all the support and affection she gave me.*

## *Curriculum Vitae*

### **Maxime Melikian**



French Nationality

Born on February, 15<sup>th</sup> 1982 in Cannes (France)

Corresponding address :

Magnetic Resonance Centre (CERM)  
Via Luigi Sacconi 6,  
Università degli studi di Firenze  
50019 Sesto Fiorentino (FI), Italia


Phone : +39 055 457 4244

E-mail : [melikian@cerm.unifi.it](mailto:melikian@cerm.unifi.it)

Permanent address :

983b route de Tuteigny  
01170 Gex, France

Phone : +33 450 99 1636

 + 39 33 45 47 81 20

E-mail : [maxime.melikian@laposte.net](mailto:maxime.melikian@laposte.net)

#### ACADEMIC RECORD :

- 2006 **Master of biochemistry** from Université Lyon I (France). Six month stage with Pr Claude-Yves Cuilleron ; work on organic synthesis.  
Thesis title : Synthesis of P-glycoprotein inhibitors from biliary acids
- 2004 **Bachelor of Biochemistry** from Université Lyon I (France).

#### ATTENDED MEETINGS :

- 2009 **Gordon Conference** on MMPs, Les Diablerets, Switzerland
- 2008 **EU-NMR & CA** Annual Meeting, Frankfurt, Germany  
**EMBO** Course on NMR spectroscopy – il Ciocco, Italy
- 2007 Advances and Management of NMR in Life Sciences, Firenze, Italy  
XI Chianti Workshop on Magnetic Resonance, Vallombrosa, Italy

#### PUBLICATIONS :

- 2009 Bertini I, Fragai M, Luchinat C, Melikian M, Venturi C. Characterisation of the MMP-12-elastin adduct. *Chemistry*. **15** (32):7842-5. (2009).
- Bertini I, Fragai M, Luchinat C, Melikian M, Mylonas E, Sarti N, Svergun DI. Interdomain flexibility in full-length matrix metalloproteinase-1 (MMP-1). *J Biol Chem*. **284** (19) : 12821-8 (2009).
- 2008 Bertini I, Calderone V, Fragai M, Jaiswal R, Luchinat C, Melikian M, Mylonas E, Svergun DI. Evidence of reciprocal reorientation of the catalytic and hemopexin-like domains of full-length MMP-12. *J Am Chem Soc*. **130** (22) : 7011-21. (2008).

## EXPERIENCE :

Since 2006 **NMR spectroscopy** of proteins : monodimensional, HSQC, Trosy, HNCO, HNCA, CBCACONH, HNCACB and T1, T2, NOE mobility experiments.

**Bioinformatics** : Protein assignment, structure calculation of proteins and docking calculation.

In 2006 **Organic chemistry** : synthesis of P-glycoprotein inhibitors

## FOREIGN LANGUAGES :

**English** : read, spoken, written – conversation level

**Italian** : read, spoken, written – usual speaker

**Spanish** : read, spoken, written – conversation level

## SOCIAL SKILLS :

Motivated, logical thinking, serious, precise, enthusiastic, creative, scientific curiosity, team spirit.

## INTERESTS :

Developing my knowledge, traveling and going to the country-side.

Sport : Football, running, cycling.

Culture : Music, cinema, reading, cuisine.