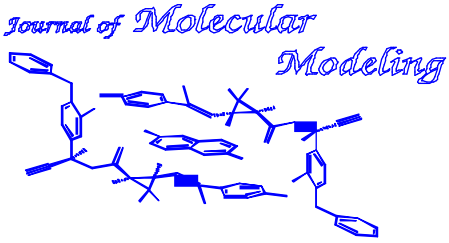
J. Mol. Model. **1997**, *3*, 17 – 30



Structural Insights into the Catalytic Domains of Human Matrix Metalloprotease-2 and Human Matrix Metalloprotease-9: Implications for Substrate Specificities

Irina Massova, Rafael Fridman,§ and Shahriar Mobashery*

Department of Chemistry and the Barabara Ann Karmanos Cancer Institute, Wayne State University, Detroit, Michigan 48202, USA; Phone: 313-577-3924; Fax: 313-577-8822 (som@mobashery.chem.wayne.edu)

§ Department of Pathology

Received: 8 October 1996 / Accepted: 24 December 1996 / Published: 20 January 1997

Abstract

Structural information for the gelatinases A (MMP-2) and B (MMP-9), two members of the matrix metalloprotease (MMP) family of enzymes, has been elusive. For the first time, computational structures for the catalytic domains of MMP-2 and MMP-9 are reported herein using the program COMPOSER and the reported three-dimensional structures of the fibroblast collagenase (MMP-1), neutrophil collagenase (MMP-8) and stromelysin-1 (MMP-3). The details of the structures of the catalytic domains of gelatinases and interactions with the protein substrate are discussed. The first analysis of the extent of hydrophobicity of surfaces in the active sites of six MMPs (including the two gelatinases reported herein) is presented to provide distinction for substrate specificity among these metalloproteases. The information from the extent of hydrophobicity/hydrophilicity analysis and general topology for these MMPs was utilized in the proposal of a method for categorization of MMPs of known three-dimensional fold. These efforts provide the first information useful to experimentalists working on the biochemical properties of these important members of the MMP family of enzymes, and provide for an opportunity to compare and contrast structures of gelatinases, collagenases and stromelysins.

Keywords: Matrix metalloprotease, gelatinase A, gelatinase B, catalytic domain structure, substrate specificity

Introduction

Members of the family of matrix metalloproteases (MMPs) play central roles in the remodeling and turnover of extracellular matrix (ECM) in normal and pathological processes, including wound healing, angiogenesis, arthritis and cancer. This family of enzymes includes collagenases, stromelysins, gelatinases [1,2], and membrane-type MMPs [3, 4, 5]. All members of the MMP family share certain common features, which include an N-terminal propeptide which maintains latency in the zymogenic forms, a zinc-containing catalytic domain, and a C-terminal fragment referred to as the hemopexin/vitronectin-like domain [1, 2]. The gelatinases contain an additional fibronectin-like region intercalated within the catalytic domain [6, 7]. The membrane-type MMPs

* *To whom correspondence should be addressed*

contain an additional transmembrane domain which plays a role in anchoring the enzyme on the plasma membrane.

Structural information for the catalytic domains of two collagenases (MMP-1 [8] and MMP-8 [9]), and stromelysin-1 (MMP-3 [10]) are now available [11]. The general folding patterns of the catalytic domains in these proteins is well conserved, and each contains two zinc- and at least one calcium-binding sites. To date, any such structural information is lacking for gelatinases. However, from amino-acid sequence alignments for MMP-2 and MMP-9, the disjointed (*vide infra*) active-site domains of these enzymes are known to contain approximately 160-168 amino-acid residues.

The catalytic domains of gelatinases, which includes the catalytic zinc-binding site, are homologous with high degree of amino acid sequence identity to those of the fibroblast collagenase (MMP-1), neutrophil collagenase (MMP-8) and stromelysin-1 (MMP-3) [1]. Gelatinases cleave gelatin, collagens type IV, V, VII, XI, fibronectin, elastin, laminin, entactin, ß-amyloid, galectin-3 and proteoglycans [1, 2, 6, 7, 12 – 16], but gelatinases have never been shown to have any activity against connective tissue collagens. It was widely believed that while MMP-1 could degrade native fibrillar collagen type I, gelatinases were only active against denatured collagen. However, Aimes and Quigley [17] have shown recently that MMP-2 can cleave collagen fibrils and native type I collagen at a Gly-Ile/Leu peptide bond, displaying a specificity and maximal rate similar to that of MMP-1. These findings indicate that gelatinases share both sequence homology and active-site topology with collagenases.

Information on the structures of gelatinases has not been available in part due to the fact that these enzymes have not been crystallized. Furthermore, since the catalytic domains of gelatinases is larger due to the insertion of a 174-175 amino acid fibronectin-like domain in the vicinity of the active site in each case, the analysis of the structure by NMR has not been attempted. We describe herein the nature of our computational models for the catalytic domains of MMP-2 and MMP-9, which should prove useful in the study of the mechanism of the catalytic function of these important enzymes. Furthermore, these structures would provide an opportunity for structure-based efforts in design of inhibitors and study of substrate preference for this family of enzymes.

Materials and Methods

The amino-acid sequences of human MMP-2 (accession no. P08253) and MMP-9 (accession no. P14780) were obtained from the GenBank. The coordinates for homologous proteins with reported structures were taken from the Brookhaven Protein Database (1cgl, MMP-1; 1mnc, MMP-8; 2str, MMP-3). The amino-acid numbering system for the MMP-1 [8], the MMP-8 [9] and MMP-3 [10] have been described previously.

The folded structures for MMP-2 and MMP-9 were determined by the use of the software COMPOSER [18] implemented in SYBYL, version 6.22 [19]. COMPOSER aligns protein amino-acid sequence for a protein of unknown structure with those of homologous proteins of known structure based on topological similarities of their secondary structure elements and amino-acid sequence identities. This protocol is used to construct the three-dimensional framework of the structure. The use of COMPOSER is recommended for proteins that have amino-acid sequence identities which exceed 30% [20]. The amino-acid identity for MMPs within the catalytic domain (which includes the active-site zinc binding fold) ranges between 56-64%. Chothia and Lesk [21] indicate that reliability of the predicted structure exceeds 90% when amino-acid sequence identities are >50%. So, the reliability of the models reported here is very high. We added 116 water molecules to the gelatinase models which occupy the average positions in the homologous proteins of known structures and the energy of the entire complex was minimized by the AMBER force field [22]. Additional force-field parameters were developed for the zinc ion[23], based on our survey of all zinc-containing metalloproteases available from the Brookhaven Protein Data Bank. The reliability of these force-field parameters for zinc were verified by energy minimization of the collagenase structure (1cge) using these parameters. For these efforts, the structure of the MMP-1 was energy-minimized by AMBER using the additional parameters for the zinc ion developed by us and the resultant structure was compared with the crystallographic coordinates. The rms deviation for the Cα for the energy-minimized and crystallographic structures was 0.51 Å, and for active-site residues, as well as residues coordinating the catalytic zinc ion was 0.33 Å. For a crystal structure with resolution of 1.9 Å, as is the case for 1cge, the reliability of the positions of atoms after refinement is ±0.2 to ±0.3 Å [24], which fall approximately in the range of the rms deviation that we see for the protein after energy minimization, using our force-field parameters.

Gelatinases have a fibronectin-like domain of 174-175 amino-acid residues, which is inserted in the primary structure of the catalytic domain. This domain is replaced by a short loop of four amino acids in collagenases and stromelysin-1. In our model building for the gelatinases, we left out the sequence for the fibronectin-like domain, hence the models are comprised of two continuous sequences of polypeptides. This omission of the fibronectin-like domain, in our opinion, would not introduce structural variation into the conformation of the active site in light of the fact that the beginning of the domain is 16-19 Å away from the catalytic zinc ion, such that it would exist laterally to the catalytic domain. This omission of the fibronectin-like domain created a gap in the primary structure; however, in light of the stability of the structure of the folded proteins, it is significant that energy-minimization of the complexes of the sequences of the two polypeptides did not move apart during the energy-minimization procedure.

The autocatalytic cleavage site of MMP-2 (...DVANYNFF...) was fitted as a substrate into the active

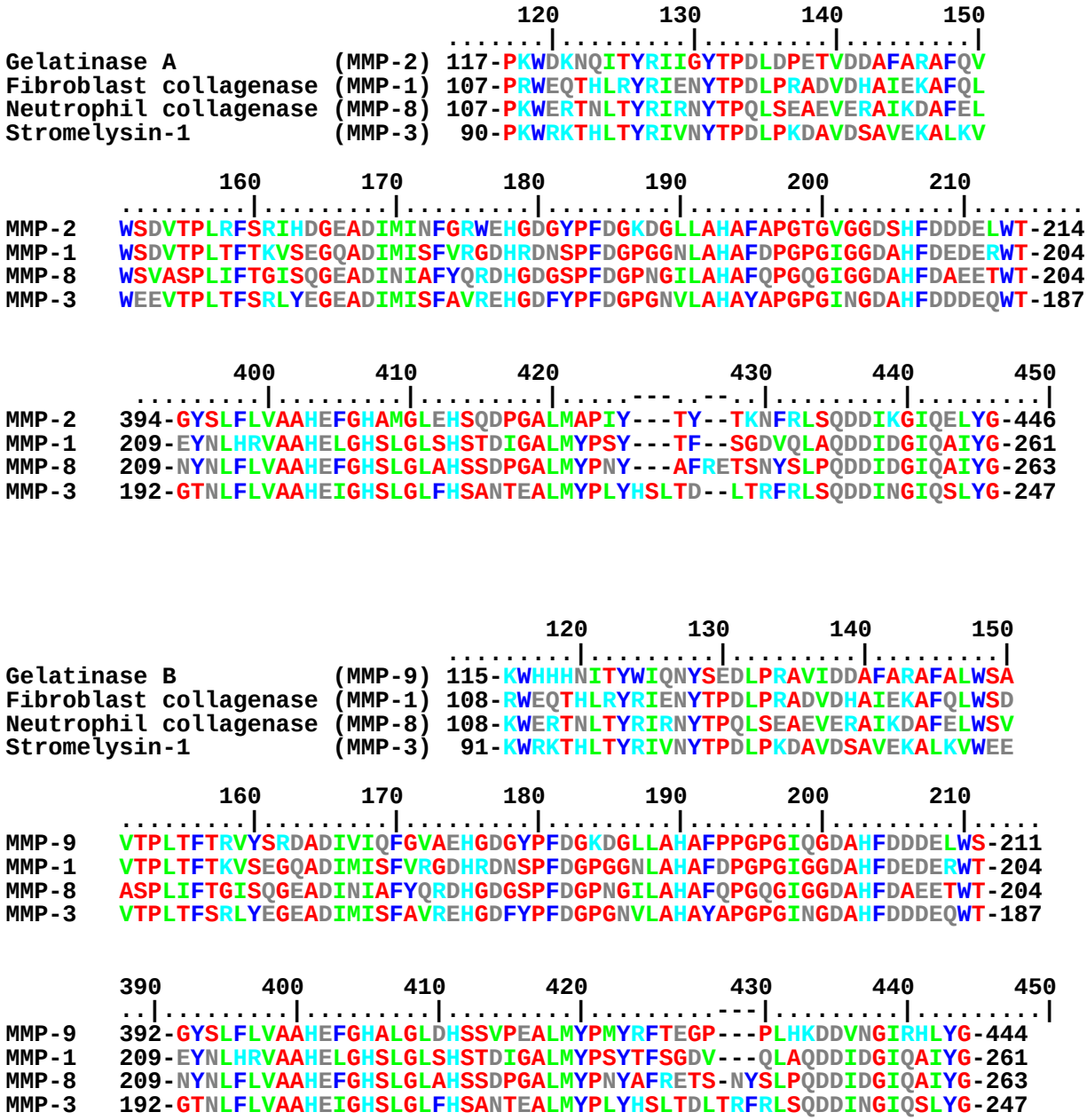


Figure 1. *Multiple amino-acid sequence alignment for the catalytic domains of human gelatinases A (MMP-2) and B (MMP-9) with human fibroblast collagenase (MMP-1), human neutrophil collagenase (MMP-8), and human stromelysin-1 (MMP-3) according to both sequence homology and topological similarity carried out by the program COMPOSER. Amino acids are colored according to their physico-chemical properties as the following: Pro, Gly, Ala, Ser, and Thr are in red; His, Arg, and Lys are in cyan; Tyr, Phe, and Trp are in blue; Asn, Asp, Gln, and Glu are in gray; Ile, Leu, Val, and Met are in green.*

site of the catalytic domain of MMP-2. The scissile carbonyl group was presented as a hydrated amide to mimic the transition state for the hydrolytic reaction. Water molecules were added to cover the substrate within the active site and the energy of the entire non-covalent enzyme-substrate complex was minimized by the AMBER force field, supplemented with our zinc parameters. The energy minimization was continued until one of these conditions was observed: (1) 50000 cycles of energy minimization were completed, (2) norm of the gradient of the energy became less than 0.001 kcal/(mol·Å), or (3) the difference in energy values for successive iterations in energy minimization cycles was within 0.001 kcal/mol.

Table 1. *The extent of sequence identity and similarity scores for MMP-2 and MMP-9 compared to MMPs with known structures.*

	MMP-2 (Pro117-Thr214...Gly394-Gly446)			MMP-9 (Lys115-Ser211...Gly392-Gly444)		
Enzyme	Homology Regions	% Sequence Identity	Similarity Score	Homology Regions	% Sequence Identity	Similarity Score
MMP-1	Pro107-Thr204 Glu209-Gly261	58	30	Arg108-Thr204 Glu209-Gly261	56	36
MMP-3	Pro90-Thr187 Gly192-Gly247	64	45	Lys91-Thr187 Gly192-Gly247	59	28
MMP-8	Pro107-Thr204 Gln209-Gly263	60	22	Lys108-Thr204 Gln209-Gly263	57	27

Results and Discussion

The catalytic domains of MMP-2 and MMP-9 both share considerable sequence similarity to MMP-1, MMP-3 and MMP-8. Figure 1 shows the multiple sequence alignment of both gelatinases with the three MMPs for which structures are known, carried out by the program COMPOSER. Table 1 summarizes the extent of amino-acid sequence identities and the similarity scores among the gelatinases and the three MMPs. The similarity score is an important factor for analysis by the COMPOSER program. It is the mean difference between the amino-acid sequence identity and the identity measured after the amino-acid sequences of the two proteins that are being compared to each other have been randomized 25 times. In order for two proteins to be considered homologous, the similarity score should be higher than 3 for most cases. The higher this number, the more significant is the percent of homology for the two proteins which are being compared. For our case, this similarity score is at the minimum 22, indicating the high homology of our proteins of interest to those for which structural information is available. Percent of amino-acid sequence identity among these proteins ranges from 56% to 64%, which gives a reliability for the computed coordinates of our models within the topologically identical regions in excess of 90% [21].

We decided to carry out an additional control experiment to investigate the reliability of the COMPOSER protocol for our application. We attempted to predict the folded structure for the MMP-1 based on the known structures of the MMP-3 and MMP-8, and amino-acid sequences of the proteins. Of the several crystal structures of the MMP-1 which are already available [8, 25], we used the structure 1cge for comparison with its predicted structure. The extent of sequence identity of the human MMP-1 with the MMP-3 and MMP-8 are 60% and 64%, and the similarity scores are 28 and 30, respectively. Figure 2 shows the outcome of the COMPOSER analysis for this protein and compares the predicted structure with that reported for the crystal structure. With the sole exception of the loop comprised of residues 241-247 (at 5 o'clock in Figure 2), the remainder of the backbone elements of the predicted structure superimposed nearly perfectly on the crystal structure. The similarity of the two structures is remarkable since no energy-minimization was carried out in this case. Indeed, the position of the loop may improve if minimization of the energy was carried out. If the loop is left out of the analysis, the rms deviation for the Cα from residues 112-261 for the predicted and X-ray structures is 0.46 Å; that for residues Arg-214, Val-215, Glu-219, His-218, His-222, and His-228, the histidines are the ligands to the active-site zinc ion and the other three are important active-site residues, for this protein for the two structures was 0.24 Å. The same analysis was carried out in predicting the structure of the MMP-8, and the predicted structure was similar to the crystal structure in this case as well (data not shown).

These control computations demonstrated the utility of the COMPOSER software in accurately predicting the structure of the catalytic domains of MMPs. Therefore, this program was used for the generation of the structures of the catalytic domains for MMP-2 and MMP-9. The predicted structures for the backbones of the structurally conserved regions for the two gelatinases superimposed well with those for the three known structures for MMPs (data not shown). Therefore, the general fold is extremely well-conserved and the differences in the structures of the catalytic domains of the two gelatinases and other MMPs rest with the variations on the individual amino-acid substitutions within the sequences of the proteins (*vide infra*). Figure 3 shows the energy-minimized catalytic domain for MMP-2 (that for MMP-9 is highly similar). Table 2 shows the major secondary elements of the structures of these two gelatinases.

The two catalytic domains for MMP-2 and MMP-9 are not only homologous to other MMPs of known structure, but

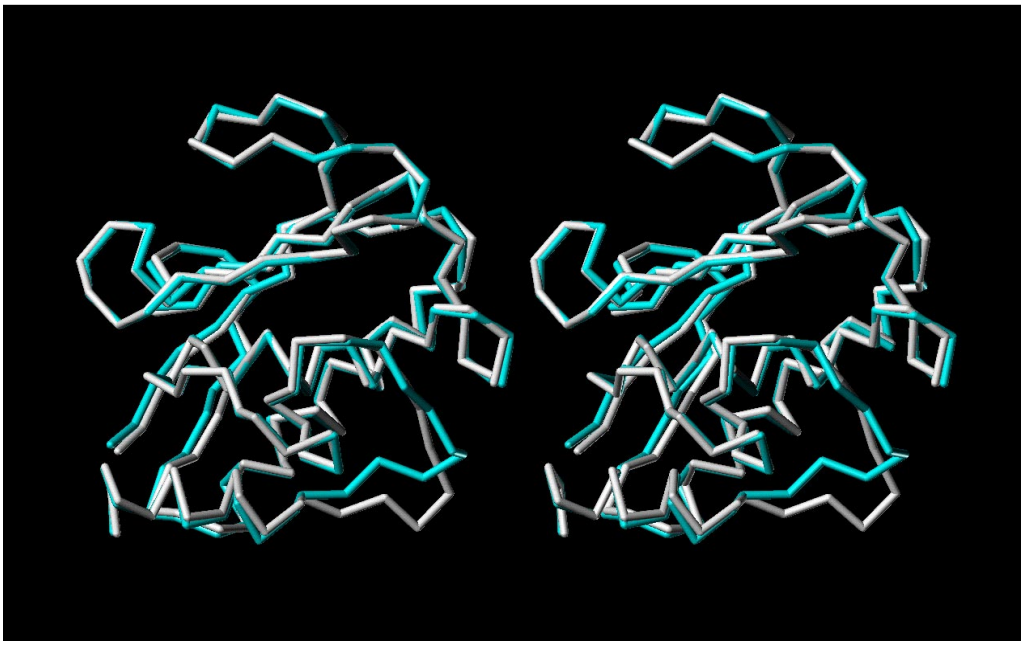


Figure 2. *Superimposition of the backbone of the predicted structure (in white) and the crystal structure (in cyan) for the MMP-1.*

they are highly homologous to one another. The percent of identity for the entire sequences of MMP-2 and MMP-9 is 48%, and when this comparison is made for the catalytic domains, the numbers are even higher. One sees from Lys-118 to Thr-214 and from Gly-394 to Gly-446 for MMP-2, and from Lys-115 to Ser-211 and from Gly-392 to Gly-444 for MMP-9 an identity of 65% [26]. Interestingly, the highly conserved regions include the zinc- and calcium-binding loops, an α-helix (B) and a β-strand (IV) within the active site, and the S_1'-binding pocket [27]. It is noteworthy that all amino-acid residues which are within the binding region of the substrate are strictly conserved, and only few residues pointing to the outside of the binding site are variable. The only variable residues within the active site are Glu-412, Ala-422, Ile-424 for MMP-2 and Asp-410, Tyr-420, Met-422 for MMP-9, respectively. It is important to point out that for two α-helices (A and C), the amino-acid residues which are located on the surface of the protein core are not conserved, however, those which point inward are highly conserved. Furthermore, the amino-acid residues sequestered on turns and loops tend to be more variable.

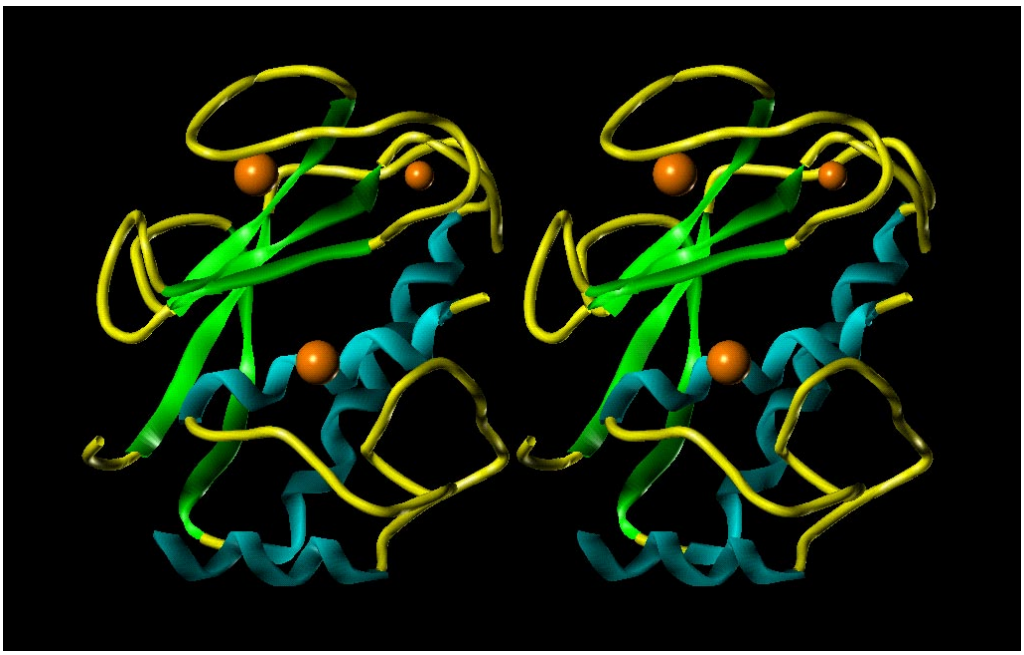


Figure 3. *The stereoview of the energy-minimized structure for the catalytic domain of MMP-2. The α-helices are in blue, the β-strands are in green, and other secondary structures are shown in yellow. The smaller orange sphere (at 1 o'clock) is the calcium ion, and the two larger spheres are zinc ions.*

Table 2. *Summary of the secondary structure elements of MMP-2 and MMP-9.*

MMP-2		MMP-9	
Segment [a]	**Structure [b]**	**Segment [a]**	**Structure [b]**
Gln123-Ile128	β-Strand I	Asn120-Ile125	β-Strand I
Pro137-Val154	α-Helix A	Arg134-Val151	α-Helix A
Pro156-His163	β-Strand II	Pro153-Tyr160	β-Strand II
Ile169-Arg175	β-Strand III	Ile166-Val172	β-Strand III
Ala192-Phe195	β-Strand IV	Ala189-Phe192	β-Strand IV
Gly203-Asp208	β-Strand V	Gly200-Asp205	β-Strand V
Leu397-Met409	α-Helix B	Leu395-Leu407	α-Helix B
Gln435-Tyr445	α-Helix C	Lys433-Tyr443	α-Helix C

[a] Numbering of the amino acids in the proteins is according to the GenBank.

[b] The terminology for the secondary structural elements is according to Lovejoy, B.; Cleasby, A.; Hassell, A. M.; Longley, K.; Luther, M. A.; Weigl, D.; McGeehan, G.; McElroy, A. B.; Drewry, D.; Lambert, M. H.; Jordan, S. R. *Science* **1994**, *263*, 375.

The substrate specificity of the gelatinases remains elusive, in part due to the fact that not many cleavage sites in protein substrates for gelatinases are known. Table 3 summarizes the known non-collagenic cleavage sites for protein substrates for gelatinases. An inspection of the amino acids around the cleavage sites indicated the preponderance of hydrophobic residues, although hydrophilic amino-acid residues are also found distributed throughout the sequences. To gain insight into substrate binding in the active site of the catalytic domain, we have modeled one sequence for the autocatalytic cleavage site of proMMP-2 into the active site of MMP-2 (Figure 4). The scissile carbonyl is shown hydrated in Figure 4 to mimic the transition-state species for the substrate hydrolysis. Figure 5 shows the schematic arrangement of interactions of the substrate in the active site. Since the active sites of the gelatinases are very similar in topology, only the active site of MMP-2 will be discussed. The active site of MMP-2 is an extended cleft region with an α-helix (B) and a β-strand (IV) forming two of the walls of the cleft. It would appear that the β-strand facilitates the orientation of the substrate for proper active-site binding by formation of an antiparallel β-sheet with the substrate. The same observation was recently made with regards to the binding of the substrate in the active site of MMP-8 [28], except that the binding of the substrate to the active site of MMP-2 appears to be somewhat more extended (discussed in more detail below).

As shown in Figure 4, many of the amino acids that line up the active-site cavity are conserved in all MMPs. However, the few exceptions are amino-acid residues 182, 188, 189, 190, 196, 395, 399, 400, 412, and 424 [26]. It is plausible that in conjunction with potential interactions of substrates with other enzyme domains, the variations of these amino acids would define the substrate specificity for each individual MMP. Furthermore, the loop comprised of residues 424-432 should be important for substrate specificity. Only residue Tyr-425 within this loop is invariant among all MMPs. This loop forms a wall of the S_1' pocket, known to be important in interactions of the substrates with the active site. The deep S_1' pocket of MMP-2 is of the same dimensions as that for the MMP-8. However, we hasten to add that the backbone of the MMP-1, in somewhat of a contrast to that for the MMP-8, traces that for MMP-2 more closely (Table 4). The principal structural difference is, however, the position of the side chain at residue Leu-399. In MMP-1 it is an arginine, which makes the S_1' cavity more shallow. Both gelatinases and MMP-8 have a leucine residue at this position that makes the cavity a deeper and more hydrophobic one.

Three hydrophobic amino-acid residues, Tyr-425, Tyr-395, and Leu-191 cluster in MMP-2 to form a hydrophobic core that constitutes the S_3' binding site. Hence, existence of a hydrophobic functionality at P_3' should be expected in preferred substrates, consistent with the results tabulated in Table 3. Of these three amino acids, that at position 395 is variable among MMPs, with substitutions of Thr, Asn, and Ile seen besides Tyr. The side chains of the substrate at positions P_2' and P_4' would point to the medium. These sites are bordered by the variable residues Leu-190, Gly-189, Asp-188, and Ile-424, which we suggest may play a role in defining substrate specificity. A tyrosine is found at position 182 of gelatinases. The orientation of this amino acid is such that it would force the substrate to bind the active site of MMP-2 as an extended β-strand. In contrast, a serine is found at this position in the MMP-1 and MMP-8. Grams et al. [28] have

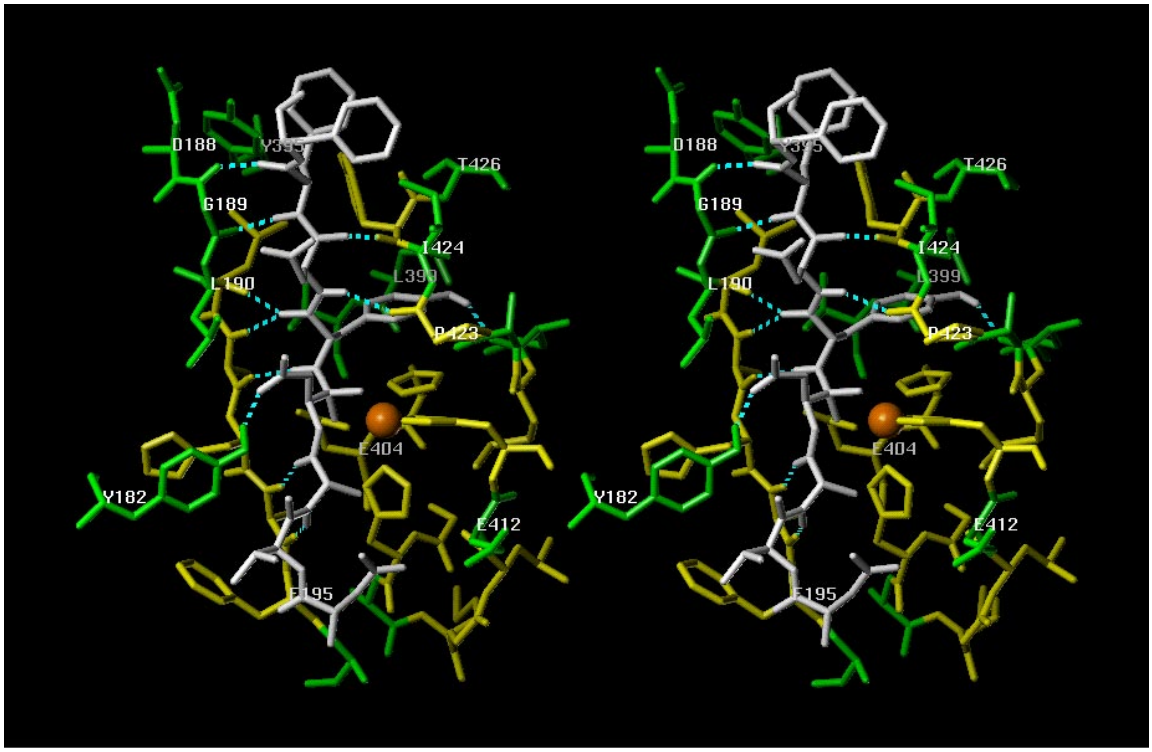


Figure 4. *Stereoview of the energy-minimized model for the complex of a peptide substrate based on an autocatalytic cleavage site of proMMP-2 in the active site of MMP-2. The peptide substrate (shown in white) is hydrated at the scissile carbonyl to mimic the transition state for the hydrolytic reaction, the oxygens of which are shown coordinated to the catalytic active-site zinc ion (orange sphere). The residues which are conserved for all MMPs are depicted in yellow, and the variable residues are in green. Only non-hydrogen atoms are shown, except when a hydrogen bond is involved between the substrate and the active site, where hydrogens are retained. Hydrogen bonds are shown in broken blue lines.*

suggested recently in their model for the binding of a substrate to the active site of the MMP-8 that the shorter side chain of this amino acid would permit a bend in the substrate conformation near this residue. Hence, according to this model, the substrate binding to MMP-8 takes place with a less extensive β-sheet formation than does for gelatinases.

The additional fibronectin-like domain of the gelatinases, which was left out from our models of the catalytic domains, has been suggested to facilitate binding of the enzyme to native and denatured collagen type I [29], but it is not directly involved in the catalytic activity [29, 30, 31,32]. A deletion mutant of the latent MMP-2 lacking this domain maintained self-activation in the presence of 4-aminophenyl

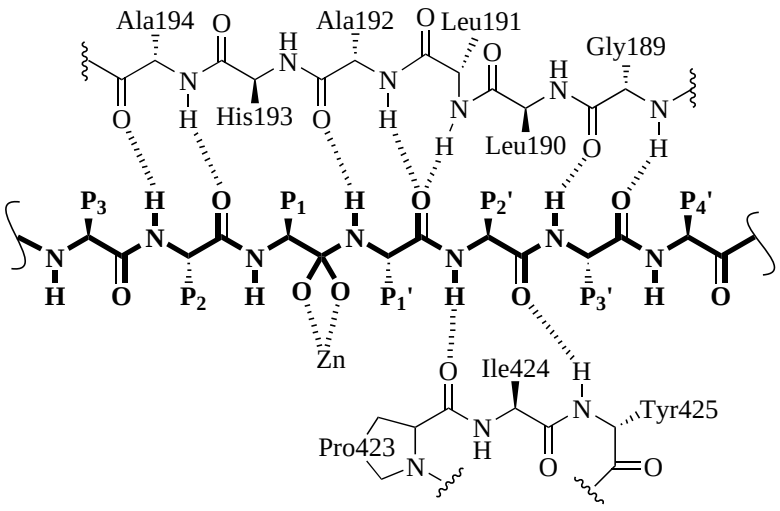


Figure 5. *Schematic representation of interactions of the autocatalytic cleavage site of proMMP-2 in the active site of MMP-2. The substrate is shown in bold-faced drawing.*

Table 3. *Some amino-acid sequences for non-collagenic protein substrates for gelatinases and the sites of enzymatic hydrolysis (indicated by ~).*

Substrate	Amino Acid Sequence	Enzyme
β-amyloid [a]	SNKGAIIGLM~VGGVVIATVI	MMP-2
β-amyloid [a]	GSNKGAIIGL~MVGGVVIATV	MMP-2
β-amyloid [a]	DSGYEVHHQK~LVFFAEDVGS	MMP-2
galectin-3 [b]	AYPGQAPPGA~YHGAPGAYPG	MMP-2 and MMP-9
aggrecan [c]	GEDFVDIPEN~FFGVGGEEDI	MMP-2 and MMP-9
FGFR-1 [d]	LEALEERPAV~MTSPLYLEII	MMP-2
MMP-2 (gelA) [e]	PRCGNPDVAN~YNFFPRKPKW	MMP-2
MMP-9 (gelB) [f]	PRCGVPDLGR~FQTFEGDLKW	MMP-2
MMP-9 (gelB) [f,g]	YRYGYTRVAE~MRGESKSLGP	MMP-2 and MMP-9
MMP-9 (gelB) [g]	GELDSATLKA~MRTPRCGVPD	MMP-9
MMP-9 (gelB) [g]	QSTLVLFPGD~LRTNLTDRQL	MMP-9
MMP-9 (gelB) [g]	ESKSLGPALL~LLQKQLSLPE	MMP-9
human cartrilage link protein [h]	YTLDHDRAIH~IQAENGPHLL	MMP-2 and MMP-9
human cartrilage link protein [h]	HIQAENGPHL~LVEAEQAKVF	MMP-2

[a] Roher, A. E.; Kasunic, T. C.; Woods, A. S.; Cotter, R. J.; Ball, M. J.; Fridman, R. *Biophys. Res. Comm.* **1994**, *205*, 1755.
[b] Ochieng, J.; Fridman, R.; Nangia-Makker, P.; Liotta, L. A.; Stetler-Stevenson, W. G.; Raz, A. *Biochemistry* **1994**, *33*, 14109.
[c] Fosang, A. J.; Neame, P. J.; Last, K.; Hardingham, T. E.; Murphy, G.; Hamilton, J. A. *J. Biol. Chem.* **1992**, *267*, 19470.
[d] Levi, E.; Fridman, R.; Miao, H.-Q.; Ma, Y.-C.; Yayon, A.; Vlodavsky, I. *Proc. Natl. Acad. Sci. U.S.A.* **1996** (in press).
[e] Stetler-Stevenson, W. G.; Krutzsch, H. C.; Wacher, M. P.; Margulies, I. M. K.; Liotta, L. A. *J. Biol. Chem.* **1989**, *264*, 1353; a number of autocatalytic sites for MMP-2 have also been reported (Bergmann, U.; Tuuttila, A.; Stetler-Stevenson, W. G.; Tryggvason, K. *Biochemistry* **1994**, *34*, 2819).
[f] Fridman, R.; Toth, M.; Peña, D.; Mobashery, S. *Cancer Res.* **1995**, *55*, 2548.
[g] Sang, Q. X; Birkedal-Hansen, H.; Van Wart, H. E. *Biochimica et Biophysica Acta-Protein Structure and Molecular Enzymology*. **1995**, *1251*, 99.
[h] Nguyen, Q.; Murphy, G.; Hughes, C. E.; Mort, J. S.; Roughley, P. J. *Biochem. J.* **1993**, *295*, 595.

mercuric acetate, and demonstrated similar activity as the full length enzyme with a peptide substrate, but showed somewhat reduced activity with gelatin as substrate [31, 33]. In a recent publication, Bányai et al. [34] suggested that the three homologous modules of the fibronectin domain form an extension to the active-site cleft of MMP-2, and therefore may be necessary for full activity of MMP-2. However, the findings of Ye et al. [33] are in conflict with the conclusion of Bányai et al. Ye et al. [33] showed that a truncated 19-kDa catalytic domain of MMP-2 lacking the fibronectin gelatin-binding domain displayed activity close to the full length enzyme, along with similar specificity toward various substrates, including gelatin. Ye et al. [33] concluded that the fibronectin domain was located remote from the catalytic domain of MMP-2, and that the removal of the fibronectin domain had no appreciable effect on either the structure or the catalytic ability of the catalytic domain. Our results reported herein support the conclusions of Ye et al.

Unlike gelatinases, the substrate specificity for other MMPs appears to be determined by the hemopexin-like domain [35 ,36]. A recent study reported the crystal structure of the porcine synovial collagenase, which includes the catalytic and hemopexin-like domains [37]. This structure shows that the hemopexin-like domain is freely hinged to the catalytic domain by a 17-residue linker, and that the hemopexin-like domain makes little contact with the catalytic domain [36, 37]. Furthermore, the crystallized full-length porcine synovial collagenase displays no structural difference when missing the hemopexin-like domain [37]. Removal of the C-terminal domain of the interstitial collagenase and

Table 4. *Categorization of matrix metalloproteases, and some features of the active sites.*

Enzyme	Group	S_1 [a]	S_1' [b]			No. of Residues for the Loop in the S_1' Pocket [c]
MMP-1	2B	N	R	L	V	7
MMP-2	1A	L	L	L	F	7
MMP-3	1A	V	L	L	F	10
MMP-7	2A	T	Y	V	F	9
MMP-8	2B	I	L	L	Y	9
MMP-9	1A	L	L	L	P	7

[a] Residues matched to the position corresponding to residue 182 in MMP-2.

[b] Residues matched to the position corresponding to residues 399, 400 and 431 in MMP-2.

[c] Number of residues between Tyr-425 and Leu-433 (numbering according to MMP-2), which constitute a portion of the wall of the S_1' pocket. All six enzymes have Tyr and Leu at these positions, respectively.

stromelysin-1 had no effect on the activity against casein, gelatin and a peptide substrate, but affected activity against native collagen [35]. Thus, similarly to the fibronectin-like domain of the gelatinases, the hemopexin-like domain of collagenase and stromelysin may play a role in substrate specificity.

The hemopexin/vitronectin-like domain of the gelatinases share high degree of amino-acid sequence identity to the same domain of the collagenases, and recent crystal structure reported by two different groups [38, 39] indicated similarity of the structures. Domain analysis revealed an important role for the C-terminal domain of the progelatinases in regulation of activity and inhibition by TIMPs. These studies showed that the C-terminal domain is (i) the major binding region for TIMPs in the proenzyme forms [40, 41, 42], (ii) it is not required for catalytic activity [40, 43], and (iii) it is required for the plasma membrane-dependent activation of proMMP-2 [43]. All these observations permit us to suggest that the missing fibronectin and hemopexin-like domains of the gelatinases should have no effect on the geometry of our models for the catalytic domains.

Finally, the zinc- and calcium-binding sites in the catalytic domains of gelatinases appear nearly identical to those in the MMPs of known structure. Figure 6 shows a schematic representation of the metal-binding sites.

Figure 7 shows the Connolly water-accessible surfaces for the active sites for six matrix metalloproteases, shown from the same perspective. At first glance, the overall size and topology of the six active sites are similar, hence one would anticipate some overlap in substrate profiles. This

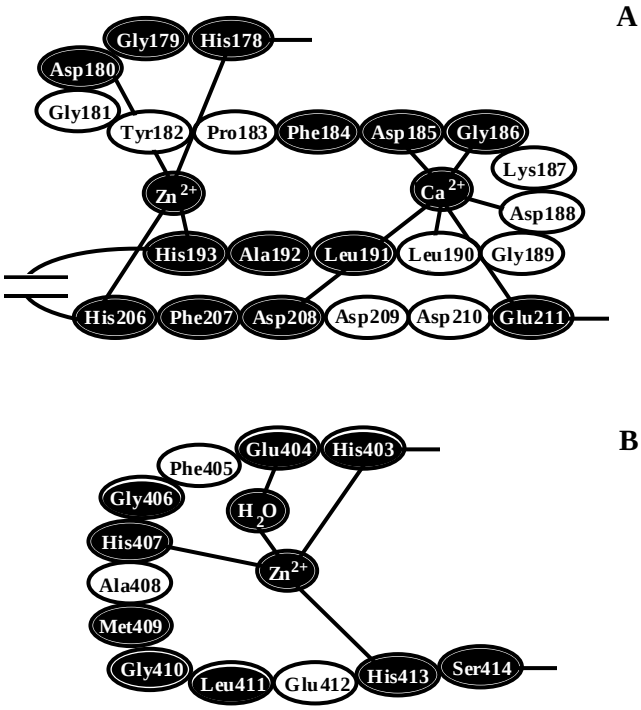


Figure 6. *Schematic of metal-binding sites in MMP-2 for (A) the structural metal-binding sites and (B) the catalytic metal-binding site. The hydrolytic water (shown in part B) is coordinated to the zinc ion, and in turn, it makes a hydrogen bond to the side chain of the active-site general base, Glu-404. Conserved residues in all MMPs are shown in black (exceptions are Asp-178 for MMP-11, Arg-179 for MMP-1, Met-411 for MMP-7, and Thr-414 for MMP-11; all other MMP have Leu-409), and residues shown in white are variable.*

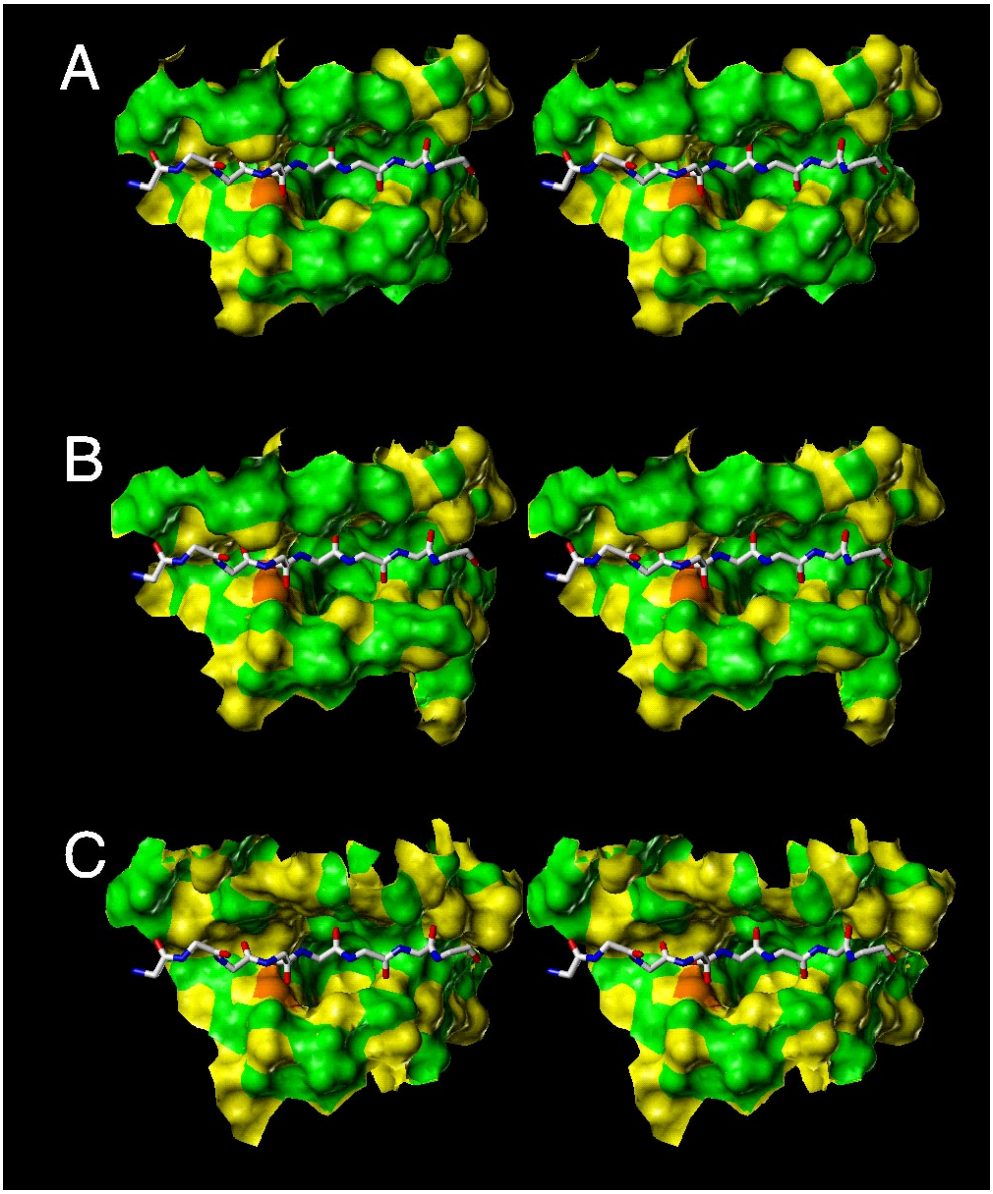


Figure 7 (continuous next page). *The stereoviews for the water-accessible Connolly surfaces for the active sites of MMPs are colored according to the following criteria: hydrophobic surfaces are shown in green, hydrophilic surfaces are in yellow, except that contributed by the zinc ion, which is depicted in orange. The water-accessible surfaces are comprised of both side-chain and main-chain functions, of which the latter is primarily involved in hydrogen bonding with substrate backbone amide functions. Only the backbone of a hypothetical substrate is shown bound in the active site, the atoms of which are colored according to atom types: white for carbon, red for oxygen, and blue for nitrogen; hydrogens are not shown. The substrate scissile carbonyl is depicted as hydrated, to mimic the structure of the transition-state species, which the enzyme stabilizes. The perspectives for figures in parts A-F are identical. The cavity in the middle of each structure is the S_1' pocket. The subsites to the right of the zinc ion are the primed sites, and those to the left are unprimed sites. (A) MMP-2 (reported in this manuscript), (B) MMP-9 (also reported in this manuscript), (C) MMP-1, (D) MMP-8, (E) MMP-3, and (F) MMP-7.*

notion finds support in recent publication of Aimes and Quigley [17], showing that MMP-2 can cleave fibrillar and native collagen type I at the same position as MMP-1, with a comparable rate. However, K_m for MMP-2 is approximately eight-fold higher than that for MMP-1, which may be due to the small differences in the two active sites. Insofar as K_m may approximate K_s, this difference may account for approximately 1 kcal/mol of binding energy difference, which is relatively small and may be the consequence of differential effects between the two active sites in hydrophobicity, steric hindrance, or even weak electrostatic interactions.

Based on the extent of surface hydrophobicity, we can divide these enzymes (shown in Figure 7) into two groups, 1

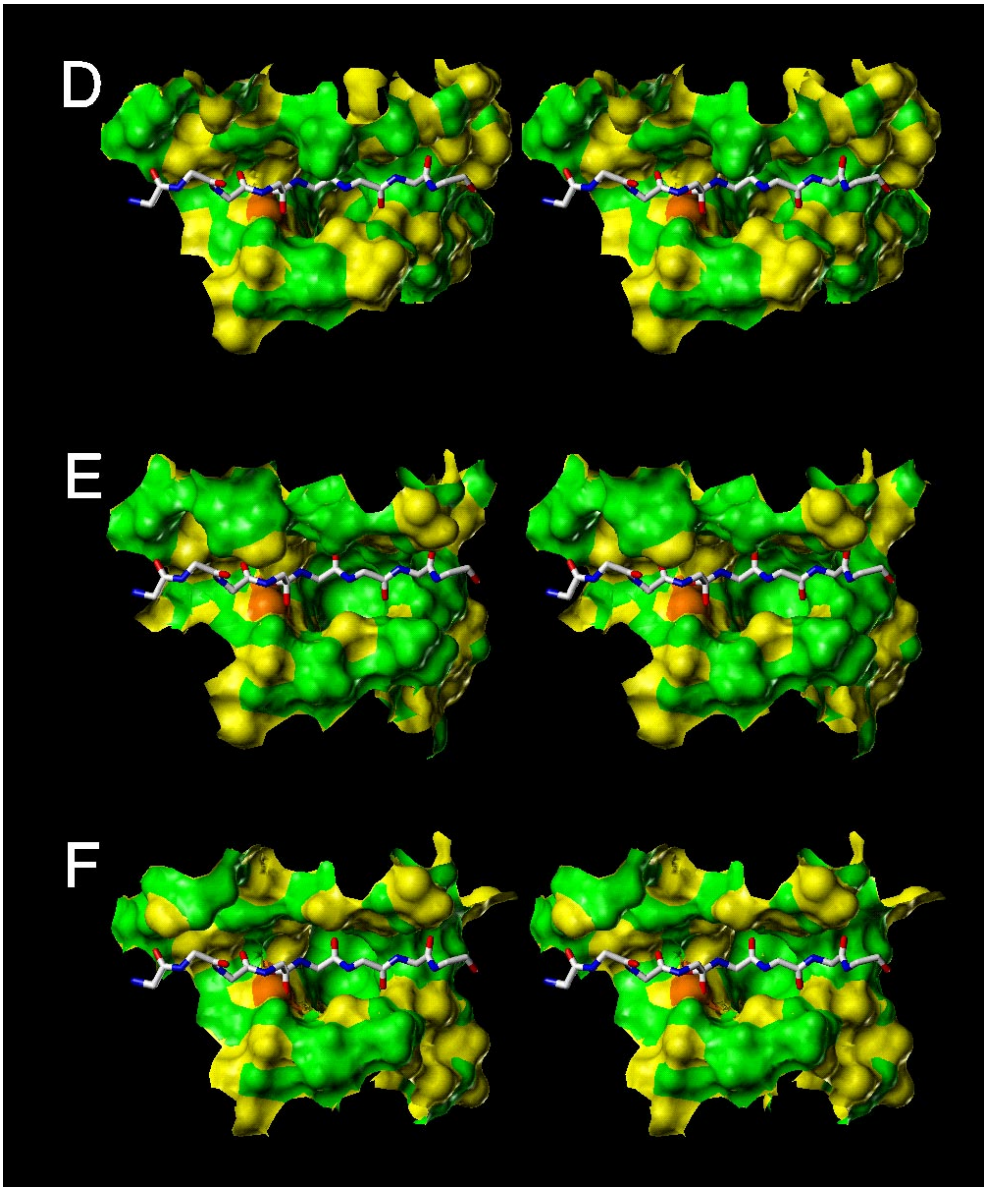


Figure 7 (continued).

and 2. Group 1 includes MMP-2, MMP-3 and MMP-9 (Figures 7A, E, and B), and group 2 includes MMP-1, MMP-7 and MMP-8 (Figures 7C, F, and D). In assessing the extent of hydrophobicity for each of these enzymes, we have analyzed the 32 amino acids that create each active site. Group 1 enzymes (MMP-2, MMP-3 and MMP-9) had 65–67% of their active sites as hydrophobic surface, in contrast to 54-57% for group 2 enzymes (MMP-1, MMP-7 and MMP-8). A portion of the hydrophilic surface in the active site is involved in electrostatic anchoring of the backbone of the substrate by hydrogen bonding, which is similarly present in all these enzymes. The remainder of the surface should play a role in defining the specificity for substrate by each of these enzymes. In our opinion the difference of 10% in the hydrophobic surfaces of the group 1 and group 2 enzymes is quite significant for the substrate specificity issue.

A second issue germane to the enzyme specificity is the degree of openness of the active site, particularly on the unprimed [27] portions. These six related enzymes can be divided into groups A and B based on this property. Group A includes MMP-2, MMP-3, MMP-7, and MMP-9, all of which have a tyrosine in the left upper quadrant of the active site (Figure 7), whereas MMP-1 and MMP-8 (belonging to group B) have the less bulky and less hydrophobic serine at this position. The bulk of Tyr-182, as well as that for a Leu-190 near the S_1 site, is the structural factor that influence the decrease of gelatinase activity with synthetic peptide substrates possessing residues larger than glycine or alanine at position P_1 [44, 45]. The classifications of MMPs of known structures by our system of groups 1 and 2, and groups A and B are given in Table 4.

Table 5. *Percentages for preferences for hydrophobic residues near the cleavage sites for known non-collagenic substrates (references are given in Table 3) for MMP-2.*

	P_5	P_4	P_3	P_2	P_1	P_1'	P_2'	P_3'	P_4'	P_5'
Autocleavage sites of MMP-2 in 2 min of incubation [a]	43	57	86	29	57	100	71	57	86	71
Other autocleavage sites of MMP-2 during 8 min to 15 h of incubation [b]	71	59	82	59	53	76	71	76	65	47
All sites [c]	58	58	88	58	55	88	67	73	73	58

[a] Total of seven. [b] Total of 17. [c] Total of 33.

The issue of the hydrophobic S_1' pocket (shown as the cavity in the center of the surfaces shown in Figure 7) has been addressed in many publications [9, 28, 46], and we will not elaborate much on this matter here. However, we note that along with the nature of the residues that line up this pocket, the length of the loop which forms the larger portion of the wall of this cavity should be important for flexibility of the enzyme to accommodate the side chain of the residue at position P_1' of the substrate. Among the six enzymes shown in Figure 7, MMP-7 has the smallest S_1' pocket, lined up with bulky residues such as valine and tyrosine (Table 4), in good agreement with the kinetics findings of Netzel-Arnett et al. [44] that MMP-7 is the least tolerant of these metalloproteases for the bulk of the residue that fits in the pocket. On the other hand, for the two collagenases, the loop is the larger for MMP-8 (9 residues; Table 4), compared to a shorter loop (7 residue) for the MMP-1. Our observation here is consistent with the kinetic measurements of Netzel-Arnett et al. [47], which indicated that a five-fold increase in k_{cat}/K_m activity for the MMP-8 when the P_1' residue in the substrate was changed from isoleucine to the bulkier tyrosine, whereas the same structural change in substrate resulted in a two fold decrease in k_{cat}/K_m for the MMP-1. The issue of the size of the S_1' pocket for gelatinases was discussed earlier in the manuscript.

Our structure-based categorization of these enzymes based on the hydrophobic/hydrophilic properties of the surfaces of the active sites, and the general topology (given in Table 4), agree closely with the classification offered for these enzymes based on specificity toward natural substrates [1]. Our groups 1A, 2B, and 2A correspond to groups I, II, and III of Woessner, respectively. The only exception is MMP-3 which falls into group 1A according to our structure-based scheme, and into group III according to Woessner. A more precise categorization of this family of enzymes should await further structural information for other related enzymes, both for the catalytic domains and entire enzymes.

From our discussion of the issue of the extent of hydrophobicity/hydrophilicity of the surfaces of the active sites for gelatinases would follow that substrate preference is not an all-or-none issue, rather the specificity preference represents a continuum. Bergmann et al. [48] observed that all autolytic sites, of which there are in excess of 20, within the primary structure of the latent form of MMP-2 are located in the N-terminal propeptide and catalytic domains, and none was found in the fibronectin or hemopexin domains. We analyzed the cleavage sites of autodegradation for MMP-2 [48] using our model for the catalytic domain of MMP-2 and the crystal structure for the N-terminal propeptide of MMP-3 [49]. Since the coordinates for two small stretches of amino acids within the propeptide of MMP-3 did not refine [49], we obtained structural information for 19 of the 23 cleavage sites. Analysis revealed that among these 19 sites, six were parts of β-strands, seven were within turns, loops or random coils, and six were located in α-helices.

The lack of this all-or-none effect in substrate preference for gelatinases may actually be an evolutionary adaptation by these enzymes. A statistical analysis of the preponderance of hydrophobicity for amino-acid residues in substrates for MMP-2, including the autolytic sites, tabulated in Table 5 reveals the clear preference of MMP-2 for hydrophobic substrates [50]. The strictest requirement seen in protein substrates is for the P_1' position, but also unexpectedly, for the P_3 position, for each of which an 88% preference for hydrophobic amino acids is observed. The preference for hydrophobicity at the P_3 position is consistent with our model of substrate binding in the active site as an extended β-strand in the unprimed direction, which forces the P_3 amino acid to occupy the hydrophobic space near the Tyr-182 residue. This portion of the active site is primarily composed of hydrophobic functionalities (upper left quadrant of Figure 7A). Because of the aforementioned "continuum" in substrate preference by gelatinases, these enzymes can accommodate a fairly diverse range of peptide substrates [44, 45]. However, the results of the statistical analysis in Table 5 reveal that the preference for substrate structure within the first couple minutes of the enzymic reaction (i.e., kinetic control), as opposed to lengthier incubation times (i.e., thermodynamic control), are quite distinct for P_5, P_2, P_3', and P_5'. In the first two minutes of the autoactivation of MMP-2 preference is seen for hydrophilic residues at P_2 (in more than two-thirds

of the cases) and virtually no discrimination for the nature of hydrophobicity/hydrophilicity at positions P_5, P_4, P_1 and P_3' (Table 5). Overall, MMP-2 prefers more hydrophobic residues at primed sites of substrate than at unprimed positions, which is consistent with the description of the active site surface depicted in Figure 7A.

Our analysis of the computational models for MMP-2 and MMP-9 provides for the first time an indepth insight into the structural parameters for the catalytic domains of gelatinases, and discusses the implication for substrate specificity. The information disclosed in this manuscript should stimulate further work on structures for these enzymes and should prove helpful in design of inhibitors and analysis of the mechanisms of the biological function in the future.

Acknowledgments. IM is a recipient of the Rumble and Heller Predoctoral Fellowships. The research in the Chemistry Department was supported in part by grants from the Karmanos Cancer Institute and the U.S. Army. The work in the Pathology Department was supported by the NIH grant CA 61986 (to RF).

References

1. Woessner, J. F., Jr. *FASEB J.* **1991**, *5*, 2146.
2. Matrisian, L. M. *Trends Genet.* **1990**, *6*, 121.
3. Sato, H.; Takino, T.; Okada, Y.; Cao, J.; Shinagawa, A.; Yamamoto, E.; Seiki, M. *Nature* **1994**, *370*, 61.
4. Takino, T.; Sato, H.; Shinagawa, A.; Seiki, M. *J. Biol. Chem.* **1995**, *270*, 23013.
5. Will, H.; Hinzmann, B. *Eur. J. Biochem.* **1995**, *231*, 602.
6. Collier, I. E.; Wilhelm, S. M.; Eisen, A. Z.; Marmer, B. L.; Grant, G. A.; Seltzer, J. L.; Kronberger, A.; He, C.; Bauer, E. A.; Goldberg, G. I. *J. Biol. Chem.* **1988**, *263*, 6579.
7. Wilhelm, S. M.; Collier, I. E.; Marmer, B. L.; Eisen, A. Z.; Grant, G. A.; Goldberg, G. I. *J. Biol. Chem.* **1989**, *264*, 17213.
8. Lovejoy, B.; Cleasby, A.; Hassell, A. M.; Longley, K.; Luther, M. A.; Weigl, D.; McGeehan, G.; McElroy, A. B.; Drewry, D.; Lambert, M. H.; Jordan, S. R. *Science* **1994**, *263*, 375.
9. Stams, T.; Spurlino, J. C.; Smith, D. L.; Wahl, R. C.; Ho, T. F.; Qoronfleh, M. W.; Banks, T. M.; Rubin, B. *Nature Struct. Biol.* **1994**, *1*, 119.
10. Gooley, P. R.; O'Connell, F.; Marcy, A. I.; Cuca, G. C.; Salowe, S. P.; Bush, B. L.; Hermes, J. D.; Esser, C. K.; Hagmann, W. K.; Springer, J. P.; Johnson, B. A. *Nature Struct. Biol.* **1994**, *1*, 111.
11. During preparation of this manuscript the crystal structure for matrilysin (MMP-7) also became available [Browner, M. F.; Smith, W. W.; Castelhano, A. L. *Biochemistry* **1995**, *34*, 6602].
12. Nguyen, Q.; Murphy, G.; Hughes, C. E.; Mort, J. S.; Roughley, P. J. *Biochem. J.* **1993**, *295*, 595.
13. Sires, U. I.; Griffin, G. L.; Broekelman, T. J.; Mecham, R. P.; Murphy, G.; Chung, A. E.; Welgus, H. G.; Senior, R. M. *J. Biol. Chem.* **1993**, *268*, 2069.
14. Fosang, A. J.; Neame, P. J.; Last, K.; Hardingham, T. E.; Murphy, G.; Hamilton, J. A. *J. Biol. Chem.* **1992**, *267*, 19470.
15. Ochieng, J.; Fridman, R.; Nangia-Makker, P.; Liotta, L. A.; Stetler-Stevenson, W. G.; Raz, A. *Biochemistry.* **1994,** *33*, 14109.
16. Roher, A. E.; Kasunic, T. C.; Woods, A. S.; Cotter, R. J.; Ball, M. J.; Fridman, R. *Biophys. Res. Comm.* **1994**, *205*, 1755.
17. Aimes, R. T.; Quigley, J. P. *J. Biol. Chem.* **1995**, *270*, 5872.
18. Blundell, T. L.; Carney, D. P.; Gardner, S.; Hayes, F. R. F.; Howlin, B.; Hubbard, T. J. P.; Overington, J. P.; Singh, D. A.; Sibanda, B. L.; Sutcliffe, M. J. *Eur. J. Biochem.* **1988**, *172*, 513.
19. Tripos Associates, 1699 S. Hanley Road, Suite 303, St Louis, MO 63144.
20. Srinivasan, N.; Blundell, T. L. *Protein Eng.* **1993**, *6*, 501.
21. Chothia, C.; Lesk, A. M. *EMBO J.* **1986**, *5*, 823.
22. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
23. Bonded model for the zinc ion was used with the following constants: force constant for bonds is 400 kcal/mol·Å; force constant for the angle X-Y-Zn is 55 kcal/mol·rad^2, where X is hydrogen, 70 kcal/mol·rad^2, where Y is nitrogen, and 100 kcal/mol·rad^2 in all other cases, including the X-Zn-Y angles. Parameters for van der Waals interactions are 1.85 Å and 0.06 kcal/mol. Equilibrium bonds and angles were averaged from the available crystallographic data from other metalloproteases.
24. Fersht, A. *Enzyme Structure and Mechanism*, 2nd Ed., **1985**, pp 5-7, W.H. Freeman and Co., New York.
25. Lovejoy, B.; Hassell, A. M.; Luther, M. A.; Weigl, D.; Jordan, S. R. *Biochemisty* **1994**, *33*, 8207.
26. Amino-acid numbering is according to the numbering for the sequences of MMP-2 and MMP-9 available from GenBank.
27. The terminology for specificity of subsites in the active site and the complementary features for the substrate are after the recommendations of Berger and Schechter [Berger, A.; Schechter, I. *Phil. Trans. R. Soc. London B* **1970**, *257*, 249].
28. Grams, F.; Reinemer, P.; Powers, J. C.; Kleine, T.; Pieper, M.; Tschesche, H.; Huber, R.; Bode, W. *Eur. J. Biochem.* **1995**, *228*, 830.
29. Steffensen, B.; Wallon, U. M.; Overall, M. *J. Biol. Chem.* **1995**, *270*, 11555.

30. Strongin, A. Y.; Collier, I. A.; Krasnov, P. A.; Genrich, L. T.; Marmer, B. L.; Goldberg, G. I. *Kidney International* **1993**, *43*, 158.
31. Murphy, G.; Nguyen, Q.; Cockett, M. I.; Atkinson, S. J.; Allan, J. A.; Knight, C. G.; Willenbrock, F.; Docherty, A. J. *J. Biol. Chem*. **1994**, *269*, 6632.
32. Allan, A. J.; Docherty, A. J. P.; Barker, P. J.; Huskissonm, N. S.; Reynolds, J. J.; Murphy, G. *Biochem. J.* **1995**, *309*, 299.
33. Ye, Q.-Z.; Johnson, L. L.; Yu, A. E.; Hupe, D. *Biochemistry* **1995,** *34*, 4702.
34. Bányai, L.; Tordai, H.; Patthy, L. *J. Biol. Chem.* **1996**, *267*, 12003.
35. Murphy, G.; Allan, J. A.; Willenbrock, F.; Cockett, M. I.; O'Connell, J. P.; Docherty, A. J. *J. Biol. Chem*. **1992**, *267*, 9612.
36. Bode, W. *Structure* **1995**, *3*, 527.
37. 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.; Blow, D. M. *Structure.* **1995**, *3*, 541.
38. Libson, A. M.; Gittis, A. G.; Collier, I. E.; Marmer, B. L.; Goldberg, G. I.; Lattman, E. E. *Nature Struct. Biol.* **1995**, *2*, 938.
39. Gohlke, U.; Gomis-Rüth, F.-X.; Crabbe, T.; Murphy, G.; Docherty, A. J. P.; Bode, W. *FEBS Lett.* **1996**, *378*, 126.
40. Fridman, R.; Fuerst, T.; Bird, R. E.; Hoyhtya, M.; Oelkuct, M.; Kraus, S.; Komarek, D.; Liotta, L.; Merman, M.; Stetler-Slevenson, G. *J. Biol. Chem.* **1992**, *267*, 15398.
41. Howard, E. W.; Banda, M. J. *J. Biol. Chem.* **1991**, *266*, 17972.
42. Goldberg, G. I.; Strongin, A. Y.; Collier, I. E.; Genrich, L. T.; Marmer, B. L. *J. Biol. Chem.* **1992**, *267*, 4583.
43. Murphy, G.; Willenbrock, F.; Ward, R. V.; Cockett, M. I.; Eaton, D.; Docherty, A. J. P. *Biochem. J.* **1992**, *283*, 637.
44. Netzel-Arnett, S.; Sang, Q.-X.; Moor, W. G. I.; Navre, M.; Birkedal-Hansen, H.; Van Wart, H. E. *Biochemistry* **1993**, *32*, 6427.
45. McGeehan, G. M.; Bickett, D. M.; Green, M.; Kassel, D.; Wiseman, J. S.; Berman, J. *J. Biol. Chem.* **1994**, *269*, 32814.
46. Bode, W.; Reinemer, P.; Huber, R.; Kleine, T.; Schnierer, S.; Tschesche, H. *EMBO J.* **1994**, *13*, 1263.
47. Netzel-Arnett, S.; Fields, G.; Birkedal-Hansen, H.; Van Wart, H. E. *J. Biol. Chem.* **1991**, *266*, 6747.
48. Bergmann, U.; Tuuttila, A.; Stetler-Stevenson, W. G.; Tryggvason, K. *Biochemistry* **1995**, *34*, 2819.
49. Becker, J. W.; Marcy, A. I.; Rokosz, L. L.; Axel, M. G.; Burbaum, J. J.; Fitzgerald, P. M. D.; Cameron, P. M.; Esser, C. K.; Hagmann, W. K.; Hermes, J. D.; Springer, J. P. *Protein Sci.* **1995**, *4*, 1966.
50. Since only seven cleavage sites are known for MMP-9, we could not carry out the statistical analysis for this enzyme as well. However, since both the structures for the active sites of the two gelatinases are nearly identical (Figs. 7A and 7B), and that the two enzymes have similar substrate profiles, we believe that our discussion of the statistical analysis for MMP-2 is also valid for MMP-9.