Crystal Structures of Recombinant 19-kDa Human Fibroblast Collagenase Complexed to Itself‡

Brett Lovejoy,* Anne M. Hassell,* Michael A. Luther,* Debra Weigl,* and Steven R. Jordan*

Glaxo Research Institute, 5 Moore Drive, Research Triangle Park, North Carolina 27709

Received January 20, 1994; Revised Manuscript Received March 30, 1994®

ABSTRACT: Collagenase is a member of the matrix metalloproteinase (MMP) family of enzymes. Aberrant regulation of this family has been implicated in pathologies such as arthritis and metastasis. Two crystal forms of the catalytic (19-kDa) domain of human fibroblast collagenase have been determined using collagenase complexed with a peptide-based inhibitor (CPLX) as a starting model [Lovejoy et al. (1994) *Science 263*, 375]. The first crystal form (CF1) contains one molecule in the asymmetric unit and has been determined at 1.9-Å resolution with an *R* factor of 19.8%. The second crystal form (CF2) contains two molecules (A and B) in the asymmetric unit and has been determined at 2.1-Å resolution with an *R* factor of 19.7%. The catalytic domain of collagenase is spherical with an active site cleft that contains a ligated catalytic zinc ion. Collagenase shares some structural homology with the bacterial zinc proteinase, thermolysin [Matthews et al. (1972) *Nature, New Biol. 238*, 37], and the crayfish digestive peptidase, astacin [Bode et al. (1992) *Nature 358*, 164]. The amino terminus (Leu 102 to Gly 105) of CF1 and CF2 molecules A and B differs from the conformation found in CPLX by bending away from the molecule and interacting with the active site cleft of symmetry-related molecules. In this alternative conformation, both the main-chain nitrogen and carbonyl oxygen of Leu 102 ligate the symmetry-related catalytic zinc. Although there are structural differences in the active site clefts of CF1, CF2, and CPLX, a number of complex-stabilizing interactions are conserved. The structure of collagenase will be useful for developing compounds that selectively inhibit individual members of the closely related matrix metalloproteinase family.

The matrix metalloproteinase family of enzymes (MMPs)[1] is responsible for cleaving extracellular matrix components such as collagen and proteoglycans. Degradation of extracellular matrix components occurs during normal connective tissue remodeling events such as embryonic development, uterine involution, bone resorption, and wound healing (Birkedal-Hansen, 1988; Woessner, 1991). However, MMP activities appear to be associated with some pathological processes as well. Elevated levels of collagenase and stromelysin-1 activity have been found in the rheumatoid synovial fluids and tissues of arthritis patients (Cawston et al., 1984; Hayakawa et al., 1991; Woolley et al., 1978). These studies and related findings suggest that inappropriate synthesis and release of MMPs is responsible for cartilage depletion in rheumatoid arthritis patients (Henderson et al., 1990). The finding of elevated levels of a third type of MMP, gelatinase, in patients with metastatic cancer has led to the proposal for long-term studies to assess whether increased levels of gelatinase are predictive of more invasive and metastatic cancers (Zucker, 1993). However, available evidence suggests that MMPs constitute just one of several matrix remodeling tools available to metastatic tumor cells (Birkedal-Hansen et al., 1993).

Members of the MMP family have broad and overlapping substrate specificities. Most MMPs can cleave gelatin, fibronectin, and type IV and V collagens. Only fibroblast (MMP-1) and neutrophil (MMP-8) collagenase can cleave collagenase types I, II, and III (Welgus et al., 1981; Hasty et al., 1987; Birkedal-Hansen et al., 1993). The specificity of fibroblast and neutrophil collagenase is significant since collagen present in the articular cartilage of the synovial joint is predominantly type II collagen (Henderson et al., 1990). Fibroblast and neutrophil collagenase cleave collagen types I, II, and III at the Gly 775–Ile (or Leu) 776 peptide bond of component α chains (Birkedal-Hansen et al., 1993). The broad substrate specificity exhibited by MMPs, in general, coupled with the current inability to effectively modulate specific MMPs *in vivo* has prevented the determination of the exact roles for each member of the MMP family in both physiological and pathological connective tissue remodeling events.

MMPs have a long propeptide that contains a cysteine which maintains enzyme latency by binding the catalytic zinc (Windsor et al., 1991). After MMPs are secreted, this propeptide is cleaved to form the mature active enzyme. The active enzyme contains a catalytic (N-terminal) domain and a C-terminal domain. The MMP catalytic domain shares the highest sequence homology. The C-terminal domain appears to be involved in substrate binding and in presenting the catalytic domain at the site of enzymatic cleavage (Murphy, 1992). Full-length human fibroblast collagenase is 469 residues long, and the mature active enzyme consists of residues 100–469. Initial failure to express homogeneous full-length collagenase led to the localization and expression of the catalytic domain (Lowry et al., 1992). The recombinant collagenase catalytic domain (19-kDa domain) consists of residues 100–269. Truncated (19-kDa) collagenase is capable of cleaving casein, gelatin, and various peptides but cannot cleave collagen (Murphy et al., 1992; Becherer et al., 1991).

‡ Coordinates have been deposited in the Brookhaven Protein Data Bank under filenames 1CGL, 1CGE, and 1CGF for collagenase crystal forms CPLX, CF1, and CF2, respectively.

® Abstract published in *Advance ACS Abstracts,* May 15, 1994.

[1] Abbreviations: MMP, matrix metalloprotease; CF1, 19-kDa collagenase crystal form 1; CF2, 19-kDa collagenase crystal form II; CPLX, 19-kDa collagenase complexed to a peptide-based inhibitor; *sym*, symmetry related; suHEP, sea urchin hatching enzyme precursor; SMEP, soybean metalloendoproteinase; TL, thermolysin; AS, astacin; MMP-1, human fibroblast collagenase; MMP-2, 92-kDa gelatinase; MMP-3, stromelysin-1; MMP-7 (PUMP, matrilysin), uterine metalloprotease; MMP-8, neutrophil collagenase; MMP-9, 92-kDa gelatinase; MMP-10, stromelysin-2; MMP-11, stromelysin-3; TIMP, tissue inhibitor of metalloproteinases. The MMP-*x* nomenclature used is that suggested by Nagase et al. (1992).

0006-2960/94/0433-8207$04.50/0

Table 1: CF1 and CF2 Data Collection and Refinement Statistics

	CF1	CF2
resolution (Å)	1.9	2.1
R_{merge}[a]	7.3	6.2
unique reflections[b]	13611	17088
completeness of data (%)	81.9	86.7
non-hydrogen atoms (no.)	1274	2548
water molecules	102	181
R factor (%)[c]	19.8	19.7
rms bonds (Å)[d]	0.009	0.011
rms angles (deg)[d]	2.10	2.06
mean B factor – protein (Å^2)	20.2 (σ = 6.4)	27.0 (σ = 9.0)
mean B factor – water (Å^2)	31.8 (σ = 7.5)	33.0 (σ = 7.6)

[a] $R_{merge} = \sum_i|\langle I_i\rangle - I_i|/\sum_i I_i$ where $\langle I_i\rangle$ is the average of I_i over all symmetry equivalents. [b] Only reflections $F/\sigma(F) \geq 1.0$ were considered measured in CF1 and CF2 data sets. [c] R-factor calculations include all $F/\sigma(F) \geq 1.0$ data from 7 Å to the resolution limit of the respective data sets. [d] rms bond lengths and angles are the respective root-mean-square deviations from ideal values.

However, in the presence of peptide substrates, 19-kDa collagenase shows the same degree of inhibition as the full-length enzyme (Becherer et al., 1991) and has been used as a model for inhibition of collagenase (Berman et al., 1992). In order to further define the collagenase active site cleft, we have determined the structure of collagenase at higher resolutions (CF1 and CF2) in the absence of the carboxyalkylamine-based inhibitor that is present in CPLX.

EXPERIMENTAL PROCEDURES

The two crystal forms of the 19-kDa collagenase are CF1 and CF2. Mass spectrometry analysis of 19-kDa samples used in crystallization trials indicated that the protein was somewhat heterogeneous, with as many as four residues (Met 100, Val 101, Leu 102, and Thr 103) missing from the amino terminus. Both CF1 and CF2 were grown at 4 °C by vapor diffusion from 2 M sodium formate, 1 mM calcium chloride, 50 μM zinc chloride, and 0.1 M Tris, pH 7–9 (Hassell et al., 1994). CF1 and CF2 belong to space groups $P4_12_12$ with unit cell constants $a = b = 72.6$ Å and $c = 75.1$ Å and $P2_1$ with unit cell constants $a = 71.0$ Å, $b = 50.5$ Å, $c = 48.0$ Å, and $\beta = 100.0°$, respectively. There is one molecule in the asymmetric unit of CF1 and two molecules (A and B) in the asymmetric unit of CF2. Preliminary CF1 diffraction data (3.3 Å) were collected on a Siemens multiwire area detector. CF1 and CF2 diffraction data were subsequently collected using larger crystals, which were slowly warmed to room temperature, on a Rigaku R-AXIS IIC imaging plate (Table 1). The X-ray source for both the Rigaku and Siemens detectors was a Siemens rotating-anode generator operating at 40 kV/80 mA, which provided graphite monochromator filtered Cu Kα radiation of 1.54-Å wavelength. Mature CF1 crystals were rod-shaped with approximate dimensions 0.3 mm × 0.3 mm × 0.8 mm. Mature CF2 crystals had approximate dimensions 0.5 mm × 0.7 mm × 0.4 mm.

CF1 Structure Determination and Refinement. The structure of CF1 was determined by molecular replacement with the program X-PLOR (Brünger, 1990) using the crystal structure of the catalytic domain of human fibroblast collagenase complexed to a carboxyalkylamine-based inhibitor inhibitor (CPLX) as a starting model. CPLX crystals have $P6_4$ symmetry with two molecules (A and B) in the asymmetric unit. CPLX has been refined to an R factor of 18.2% against 2.4-Å data (Lovejoy et al., 1994). For molecular replacement calculations, the inhibitor and water molecules were removed from molecule A of CPLX, but the two zinc ions and a calcium ion were retained. Using the preliminary CF1 data collected on the Siemens detector, a Patterson search was computed. For the search, all $F/\sigma(F) \geq 2$ CF1 data in the resolution range 10.0–3.5 Å were used, and Patterson vectors ranging from 5 to 30 Å were selected. The Patterson search failed to find a distinct solution. Consequently, a direct rotation search using a Patterson correlation (PC) target function (Brünger, 1990) was calculated. In this search, the search model was rotated every 5° about all Euler angles in the asymmetric unit of the rotation function (Rao, 1980). At each orientation sampled, the PC target function was computed using all $F/\sigma(F) \geq 2$ data from 8.0–3.3 Å. The top solution had a correlation coefficient that was 10σ higher than the average correlation coefficient for the search. Since the direct rotation function was computed at 5° intervals about all Euler angles, the top solution was subjected to PC refinement (Brünger, 1990) to further define the correct orientation of the molecule in the asymmetric unit. After PC refinement, the solution was used in a translation search against all $F/\sigma(F) \geq 2$ data in the resolution range 7–3.3 Å. The search produced a solution that was 11σ above the mean score and 3σ above the second highest score. The R factor for this solution was 44.4% (7–3.3 Å). High-resolution R-axis data (Table 1) were used in all subsequent stages of refinement.

The translation function solution was refined using simulated annealing with the program X-PLOR (Brünger et al., 1990). Temperature factor refinement resulted in atomic B factors in excess of 50 Å^2 for residues 101–105, and this segment was omitted. Subsequent inspection of an $F_o - F_c$ map revealed distinct density extending from Asn 106 to the catalytic zinc of a symmetry-related molecule (*sym* zinc 302, where *sym* denotes the symmetry-related molecule). The program FRODO (Jones, 1985) was used to build residues 102–105 into this density as well as adjust several side chains and place the second and third calcium ions (304 and 305). Subsequent electron density maps were used to place water molecules during interactive cycles of model building and refinement with X-PLOR (Brünger et al., 1990). The final model contains residues 102–263, 2 zinc ions, 3 calcium ions, and 102 water molecules. There was no density corresponding to residues 264–269, and this segment was not modeled. In addition, there was no density corresponding to the side chain of Glu 104, and it was modeled as an alanine. The final model has an R factor of 19.8% (Table 1) and all ϕ, ψ angles lie within allowed regions of the Ramachandran plot.

CF2 Structure Determination and Refinement. The structure of CF2 was determined by molecular replacement with the program X-PLOR (Brünger, 1990) using the final model of CF1 as a search model. Residues 102–105 were removed from the search model, but all calcium and zinc ions were retained. A Patterson search was calculated using the search model against all 7.0–3.0-Å data, resulting in two distinct solutions. These solutions were independently used to calculate two-dimensional (x, z) translation searches, and two solutions (corresponding to molecules A and B) that were 10σ and 7σ above the mean value of their respective searches were obtained. To determine the relative translation between molecules A and B, molecule B was used to compute a one-dimensional translation function at the four possible origins. One of these searches produced a clear solution. Rigid body refinement of the combined translation function solutions for molecules A and B of CF2 resulted in a model with an R factor of 36.0% (7.0–2.1 Å).

The model was refined using simulated annealing with the program X-PLOR (Brünger et al., 1990), resulting in a model that had an R factor of 24.8% (7.0–2.1 Å). Inspection of an initial $F_o - F_c$ map revealed distinct density extending from Asn 106 of molecules A and B to the catalytic zinc of respective

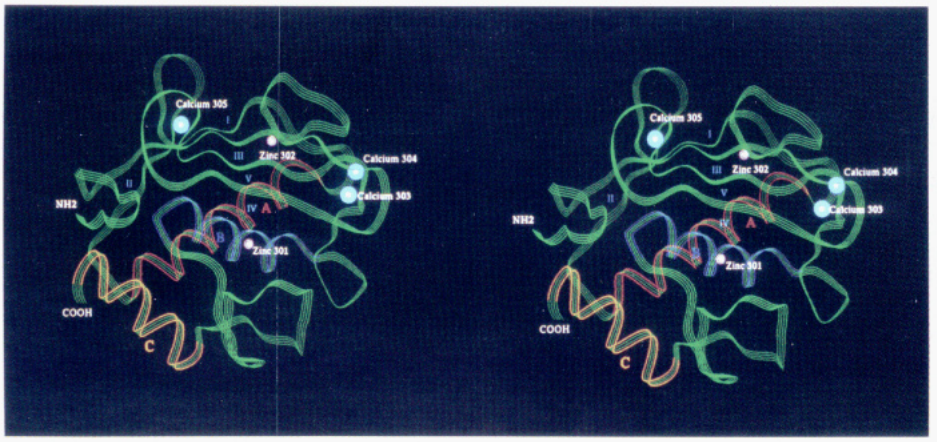

FIGURE 1: Stereo ribbon diagram of CF1. The 19-kDa collagenase contains a twisted five-stranded β sheet (strands I, II, III, IV, and V) and three α-helices (helices A, B, and C). The β strands are labeled in blue, helix A is red, helix B is blue, and helix C is orange. There are two zinc ions (purple spheres), the catalytic zinc (301) and a secondary zinc (302). In addition, there are three calcium ions (blue spheres). Calcium 303 is present in CPLX, but calcium 304 and calcium 305 are unique to CF1 and CF2. The additional calcium sites could be caused by the higher calcium concentrations present in the CF1 and CF2 crystallization conditions.

symmetry mates. The program FRODO (Jones, 1985) was used to build residues 102–105 of molecules A and B and adjust several side chains. During subsequent cycles of refinement, water molecules were identified. The final model contains residues 102–263 of molecules A and B, 2 zinc ions and 3 calcium ions associated with each molecule, and a total of 181 water molecules. The model has an R factor of 19.8% (Table 1) and all ϕ, ψ angles lie within allowed regions of the Ramachandran plot.

RESULTS

Description of the Structure. As described previously (Lovejoy et al., 1994), recombinant 19-kDa human fibroblast collagenase is a spherical molecule that contains a twisted five-stranded β sheet and three α helices (helices A, B, and C). The β sheet contains four parallel strands (I, II, III, and V) and one antiparallel strand (IV) (Figure 1). The active site cleft is bordered by β strand IV, helix B, and a stretch of random coil adjacent to the COOH terminus of helix B. The catalytic zinc (301) is at the bottom of the cleft and is ligated by His 218, His 222, and His 228. In addition to the catalytic zinc, there is a second zinc ion that interacts with an extended loop between β strands III and IV. CF1 and CF2 contain the calcium ion present in CPLX (303) as well as two additional calcium ions (304 and 305). Calcium 303 and calcium 304 are only 10.0 Å apart. Table 2 lists the ligating interactions between the zinc and calcium ions and collagenase.

Density corresponding to Met 100 and Val 101 is absent in CF1. Mass spectrometry analysis of dissolved CF1 crystals confirmed the absence of these two residues and identified Leu 102 as the amino terminus of CF1 (Hassell et al., 1994). Density corresponding to Met 100 and Val 101 is also absent in CF2. Interestingly, the Val 101–Leu 102 peptide bond is a common autolytic cleavage site in full-length collagenase (Birkedal-Hansen et al., 1993). Apart from the amino terminus (Leu 102 to Gly 105), the rms deviations among CF1, CF2, and CPLX are very low (Table 3).

The most striking difference in CF1 and CF2 is that residues 102–105 adopt different conformations than in CPLX. In these alternative conformations, residues 102–105 occupy the active site cleft of a symmetry-related molecule (Figure 2A). Due to crystallographic symmetry, the interaction between the amino terminus of each molecule in CF1 and CF2 and the active site cleft of symmetry-related collagenase molecules results in the formation of an infinite lattice of collagenase molecules. The packing arrangement of each of the three molecules found in CF1 and CF2 (CF1 and molecules A and B of CF2) are independent of each other. Consequently, the orientation of the symmetry-related active site cleft that each amino terminus binds varies. To accommodate this variability, the amino terminus must bend to a different extent in order to form a complex in CF1 and CF2. Superposition of the three molecules suggests that Gly 105 serves as a hinge that allows the amino terminus to bind the active site cleft from different orientations (Figure 2B).

Complex between the Amino Terminus and a Symmetry-Related Molecule in CF1 and CF2. The amino termini (residues 102–105) of all three molecules present in CF1 and

Table 2: Interactions between Collagenase and Zinc or Calcium Ions (in Angstroms)[a]

		CPLX			CF2	
ion	ligand	mol A	mol B	CF1	mol A	mol B
zinc 301	His 218 Nε2	2.2	2.1	2.1	2.2	2.0
	His 222 Nε2	2.0	2.0	2.1	2.2	2.2
	His 228 Nε2	2.0	2.0	2.0	2.1	2.1
	carboxy Oε1	2.0	2.1			
	carboxy Oε2	2.6	2.7			
	sym Leu 102 N			1.8	2.1	2.1
	sym Leu 102 CO			2.1	2.1	2.1
zinc 302	His 168 Nε2	2.0	2.2	2.0	2.1	2.2
	Asp 170 Oδ2	2.2	2.1	2.0	2.1	2.1
	His 183 Nε2	2.2	2.1	2.0	2.1	2.1
	His 196 Nδ1	2.0	2.1	2.0	2.1	2.0
calcium 303	Asp 175 Oδ1	2.3	2.3	2.3	2.3	2.3
	Gly 176 CO	2.2	2.3	2.3	2.0	2.3
	Gly 178 CO	2.5	2.4	2.4	2.2	2.3
	Asn 180 CO	2.2	2.4	2.4	2.2	2.4
	Asp 198 Oδ2	2.2	2.2	2.3	2.3	2.2
	Glu 201 Oε2	2.1	2.2	2.2	2.3	2.3
calcium 304	Asp 124 Oδ1			2.5	2.5	2.3
	Asp 124 Oδ2			2.3	2.3	2.4
	Glu 199 Oε2			2.3	2.3	2.3
	Glu 199 CO			2.4	2.2	2.4
	water			2.2	2.3	2.3
	water			2.3	2.4	2.4
calcium 305	Asp 158 CO			2.3	2.3	2.3
	Gly 190 CO			2.4	2.6	2.4
	Gly 192 CO			2.3	2.3	2.4
	Asp 194 Oδ1			2.5	2.5	2.4
	water			2.2	2.2	2.2
	water			2.3	2.3	2.3

[a] The Lennard-Jones σ parameter for zinc and calcium was adjusted to 2.5 and 3.0 Å, respectively, to obtain distances found in small molecule structures. Using these parameters, CPLX has been refined to an R factor of 18.2% and has rms deviations from ideal bonds and angles of 0.011 Å and 2.18°, respectively.

Table 3: rms Comparison of CF1, CF2, and CPLX (Tabulated in Angstroms)[a]

	CPLX			CF2	
	mol A	mol B[b]	CF1	mol A	mol B
CPLX mol A		**0.34**	**0.43**	**0.47**	**0.47**
CPLX mol B[b]	0.94		**0.47**	**0.55**	**0.49**
CF1	1.18	1.22		**0.35**	**0.37**
CF2 mol A	1.16	1.18	0.90		**0.47**
CF2 mol B	1.15	1.15	0.85	0.95	

[a] Structures were superimposed, and the rms deviation was calculated using the method of Hendrickson (1979). Superpositions using all main-chain atoms (N, C_α, C) are in the upper portion of the table and are boldfaced while superpositions using all non-hydrogen atoms are in the lower portion of the table. Residues 106–263 were used in all superpositions except where noted. [b] Residues 153–155 are missing in CPLX molecule B.

CF2 are represented by strong electron density (Figure 3A). In CF1 and CF2, both the nitrogen and carbonyl oxygen of Leu 102 ligate the catalytic zinc of the symmetry-related molecule (*sym* zinc 301, where *sym* denotes the symmetry-related molecule; Table 4). Thr 103 is buried in a deep hydrophobic pocket (the P1′ pocket) that is adjacent to the symmetry-related catalytic zinc. Glu 104 and Gly 105 are in extended conformation, filling the remainder of the active site cleft (Figure 3B). Many of the interactions that stabilize the complex in CF1 and CF2 are analogous to interactions between the carboxyalkylamine-based inhibitor (Figure 3C) and collagenase in CPLX. In CPLX, the central carboxy group of the inhibitor ligates the catalytic zinc while the inhibitor P1′ Leu occupies the P1′ pocket. CPLX P2′ and P3′ are in extended conformation, filling the remainder of the active site cleft in a manner comparable to Glu 104 and Gly 105 in CF1 and CF2 (Figure 3D).

The complex between symmetry-related molecules in CF1 and CF2 is stabilized by seven intermolecular hydrogen bonds (Figure 3B, Table 4). Four of these bonds correspond to protein–inhibitor hydrogen bonds present in CPLX. In CF1, these bonds are formed by Thr 103 N and *sym* Ala 182 CO (3.3 Å), Thr 103 CO and *sym* Leu 181 N (2.9 Å), Glu 104 CO and *sym* Tyr 240 N (2.9 Å), and Gly 105 N and *sym* Gly 179 CO (2.8 Å). Table 4 lists the corresponding bonds in CPLX. In collagenase, Glu 219 represents a catalytically important residue that is conserved in all metalloproteinases (Birkedal-Hansen et al., 1993; Woessner, 1991). In CF1 and CF2, the amino terminus (Leu 102 N) as well as Thr 103 $O\gamma1$ forms hydrogen bonds with *sym* Glu 219 (Table 4). In CPLX, the main-chain nitrogen of the inhibitor P1′ group, which mimics the nitrogen of the substrate scissile bond, forms a hydrogen bond with Glu 219. The final intermolecular hydrogen bond present in CF1 and CF2 is formed by Glu 104 N and Pro 238 CO (2.8 Å in CF1). The corresponding atom in CPLX (P2′ N) is too far away from Pro 238 CO (3.8 Å in molecule A of CPLX) for significant hydrogen bonding to occur.

Comparison of the CF1, CF2, and CPLX Active Site Clefts. Superposition of CF1, CF2, and CPLX reveals some structural differences in the active site cleft. The most prominent movement is in the central portion of the loop between β strands III and IV (residues 170–173). In CPLX, this segment is "pushed up" relative to CF1 and CF2 (Figure 4A). After superposition, the distance between $C\beta$ of CF1 Ser 172 and $C\delta2$ of the CPLX P1 phenyl ring is 2.6 Å. This distance suggests that residues 170–173 of CPLX moved away from the inhibitor P1 group to avoid significant unfavorable van der Waals repulsion. The superposition of CPLX and CF1 also suggests that the P1 phenyl ring prevents the formation of a favorable hydrogen bond between the Ser 172 hydroxy and P1 carbonyl groups (Figure 4A). Another difference in the active site cleft of 19 kDa collagenase is the conformation of Asn 180. The χ_1 angle of Asn 180 is −57° (gauche$^+$) in CF1 and CF2 while in CPLX it is −171° (trans). By adopting the trans χ_1 rotamer, CPLX Asn 180 $O\delta1$ and $N\delta1$ can form hydrogen bonds with N and CO of the inhibitor P1 group (3.0 and 2.7 Å, respectively; Figure 4B). In addition, the presence of the inhibitor P2′ phenyl ring prevents the χ_1 angle of CPLX Asn 180 from adopting the gauche$^+$ conformation found in CF1 and CF2.

The side chains occupying the P1′ hydrophobic pocket in CPLX (P1′ Leu) and CF1 and CF2 (Thr 103) adopt the same rotameric conformation. However, Leu is larger than Thr and extends further into the P1′ pocket. As a result, the loop containing Pro 238 appears to have been pushed outward in CPLX (Figure 4B). After superposition, the distance between CPLX P1′ Leu $C\delta1$ and CF1 Pro 238 $C\alpha$ is only 3.2 Å. At this distance, significant van der Waals repulsion would occur. In CPLX, the actual distance between P1′ Leu $C\delta1$ and Pro 238 $C\alpha$ is 4.5 Å. A direct consequence of the shift made by CPLX Pro 238 is that a potential complex-stabilizing hydrogen bond between CPLX Pro 238 CO and P2′ N is lost (CPLX molecule A, 3.8 Å; Table 4).

There are important similarities in the active site clefts of CF1, CF2, and CPLX. The residues of the active site cleft that form comparable intermolecular hydrogen bonds in CF1, CF2, and CPLX (Ala 182, Leu 181, Gly 179, and Tyr 240) occupy nearly identical positions. In addition, Glu 219 occupies the same position and forms intermolecular hydrogen bonds in CF1, CF2, and CPLX. In all crystal forms of 19-kDa collagenase, the base of the P1′ pocket is bordered by the guanidinium group of Arg 214. A water molecule present at the bottom of the P1′ pocket forms hydrogen bonds with the guanidinium group of Arg 214 (Figure 3D).

Comparison of 19-kDa Collagenase to Related Proteins. The catalytic domain of the MMP family is characterized by high sequence homology (Table 5). The MMP sequence alignment is remarkably free of gaps. However, both 92- and 72-kDa gelatinase contain an ~175 residue insert in the catalytic domain (Woessner, 1991). The gelatinase insert consists of three tandem repeats of fibronectin type II modules that endow the enzyme with gelatin-binding properties (Collier et al., 1992). In fibroblast collagenase, the gelatinase insert would correspond to an interruption between residues 208 and 210 (Table 6). Collagenase residues 208–210 are in a highly exposed loop that connects β strand V to α helix B, far away from the active site cleft. Consequently, it is likely that the gelatinase insert does not disrupt the structure of the gelatinase catalytic domain.

MMPs are all calcium activated and dependent on zinc for enzymatic activity. The side chains that participate in ligating interactions with the catalytic zinc (301), secondary zinc (302), and calcium 303 in 19-kDa collagenase are conserved in MMPs (Table 6). Zinc 301 is essential for enzymatic activity while zinc 302 and calcium 303 appear to play an important role in maintaining the structural integrity of collagenase by stabilizing the extended loop (residues 165–181) between β strands III and IV (Lovejoy et al., 1994). The strict conservation of residues that bind to these three ions suggests that the structural basis for calcium activation and zinc-dependent enzymatic activity is conserved in all MMPs.

The general fold of 19-kDa collagenase is very similar to the zinc endopeptidases thermolysin, from *Bacillus thermoproteolyticus* (Matthews et al., 1972), and astacin, from the crayfish *Astacin astacus* L. (Bode et al., 1992). Both thermolysin and astacin contain two domains, the N-terminal

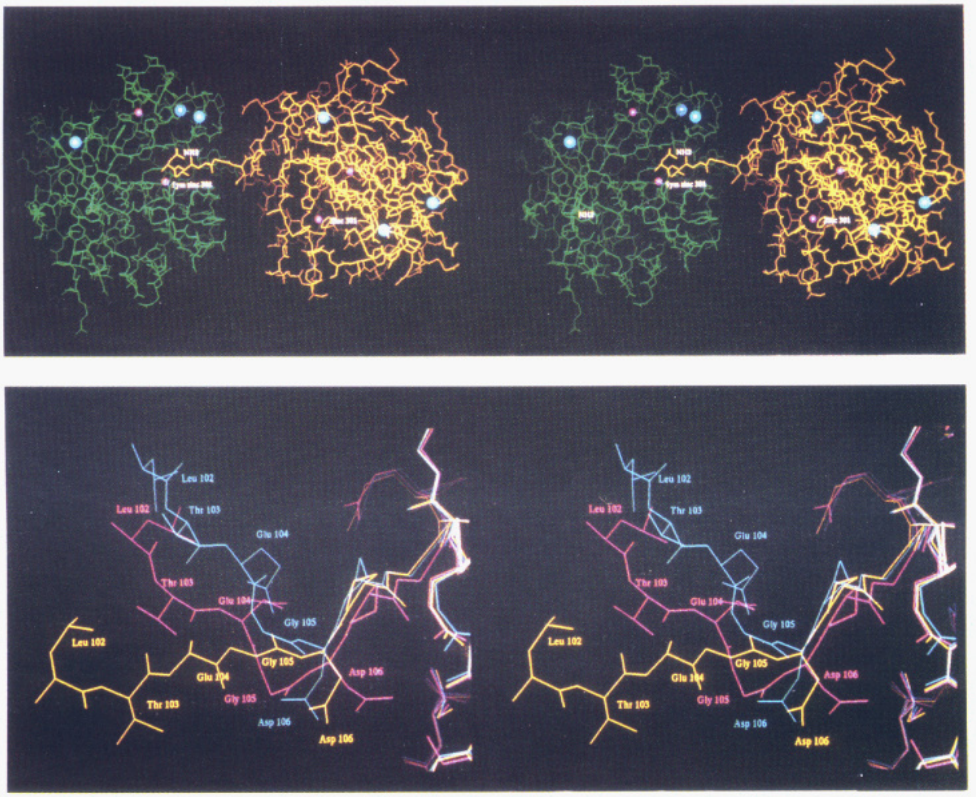

FIGURE 2: (A, top) Stereoview of the complex formed by residues 102–105 of the CF1 amino terminus (yellow) and the active site cleft of a symmetry-related molecule (green). Because of crystallographic symmetry, the amino terminus of each CF1 collagenase molecule forms a complex with the active site cleft of one symmetry-related collagenase molecule while the active site cleft of each CF1 collagenase molecule forms a complex with the amino terminus of a second symmetry-related collagenase molecule. This results in an infinite lattice of complexed collagenase molecules. CF2 molecules A and B form similar infinite complexes. (B, bottom) Stereoview of the superposition of CF1 (yellow) and CF2 molecules A (blue) and B (purple) demonstrating the various extents that the amino termini bend to form a complex. Gly 105 serves as a flexible hinge in each complex. There is no electron density corresponding to the side chain of CF1 Glu 104, and it is modeled as an alanine in CF1.

and C-terminal domains, that are separated by an active site cleft. The 19-kDa collagenase construct largely corresponds to the N-terminal domain of thermolysin and astacin. The N-terminal domain of thermolysin, astacin, and 19-kDa collagenase are defined by a twisted five-stranded β sheet and two helices that are comparable. The 19-kDa collagenase extends beyond the N-terminal domain with an extended turn followed by a C-terminal helix (helix C) that is separated from the collagenase N-terminal domain by the active site cleft (Figure 1). The structures of the astacin and thermolysin C-terminal domains are not similar to each other (Bode et al., 1992) or to the C-terminal portion of 19-kDa collagenase.

In collagenase, thermolysin, and astacin, one of the N-terminal domain helices (termed the active site helix; helix B in astacin and collagenase), the three zinc-ligating residues (His 218, His 222, and His 228 in collagenase), and the catalytic zinc share distinct structural homology (Borkakoti et al., 1994). For example, the rms difference between the Cα carbons of CF1 residues 214–224, the Cα carbon of His 228, and zinc 301 (a total of 13 atoms) to corresponding thermolysin positions (Monzingo & Matthews, 1984) (138TL to 148TL, Glu 166TL, and zinc 5TL, where TL stands for thermolysin residue numbers) is 0.76 Å. In thermolysin, astacin, and collagenase, two of the zinc-ligating residues (His 218 and His 222 in collagenase) are in the structurally conserved active site helix. Although the three thermolysin and collagenase catalytic zinc ligands are structurally conserved, the peptide chain that contains the third zinc-ligating residue does not share structural homology in collagenase and thermolysin. In collagenase, the C-terminal end of the active site helix leads into a tight turn that contains the third zinc ligand (His 228). In thermolysin, the active site helix ends with a very long loop that eventually wraps back toward the catalytic zinc and forms a helix that is unique to thermolysin. The third thermolysin zinc ligand (Glu 166TL) is in this helix. In astacin, the third zinc ligand (His 102AS, where AS stands for astacin residue numbers) is in a tight turn that, similar to collagenase, begins at the C-terminal end of the active site helix. However, after His 102AS, astacin shares no structural homology with collagenase or thermolysin.

Using inverted protein structure prediction algorithms and the structure of 19-kDa collagenase, proteins that potentially have the collagenase fold can be identified. A three-dimensional structure profile was constructed from residues 107 to 263 of CF1 (Bowie et al., 1991). This profile contains a score for introducing each of the 20 commonly occurring amino acids at positions 107–263 of CF1. Using the collagenase profile, the Genetics Computer Group database (prot_99, version 7.1, 59 091 sequences) was searched with the program PROFILESEARCH (Devereux et al., 1984). In this search, all known MMPs scored higher than any nonMMP sequence (Table 5), suggesting that 19-kDa collagenase represents the general fold of the MMP catalytic domain. The search also identified soybean metalloendoproteinase (SMEP) and sea urchin hatching enzyme precursor (suHEP). Both SMEP and suHEP have been previously identified as relatives of the MMP family (McGeehan et al., 1992; Lepage & Gache, 1990). SMEP is a zinc endoproteinase that cleaves peptide substrates with the same specificity demonstrated by MMPs and is inhibited by naturally occurring inhibitors of MMPs (McGeehan et al., 1992). Like MMPs, suHEP is a calcium-activated, zinc-dependent protease that is involved in extracellular matrix degradation (Lepage & Gache, 1990). The high scores for SMEP and suHEP suggest that these

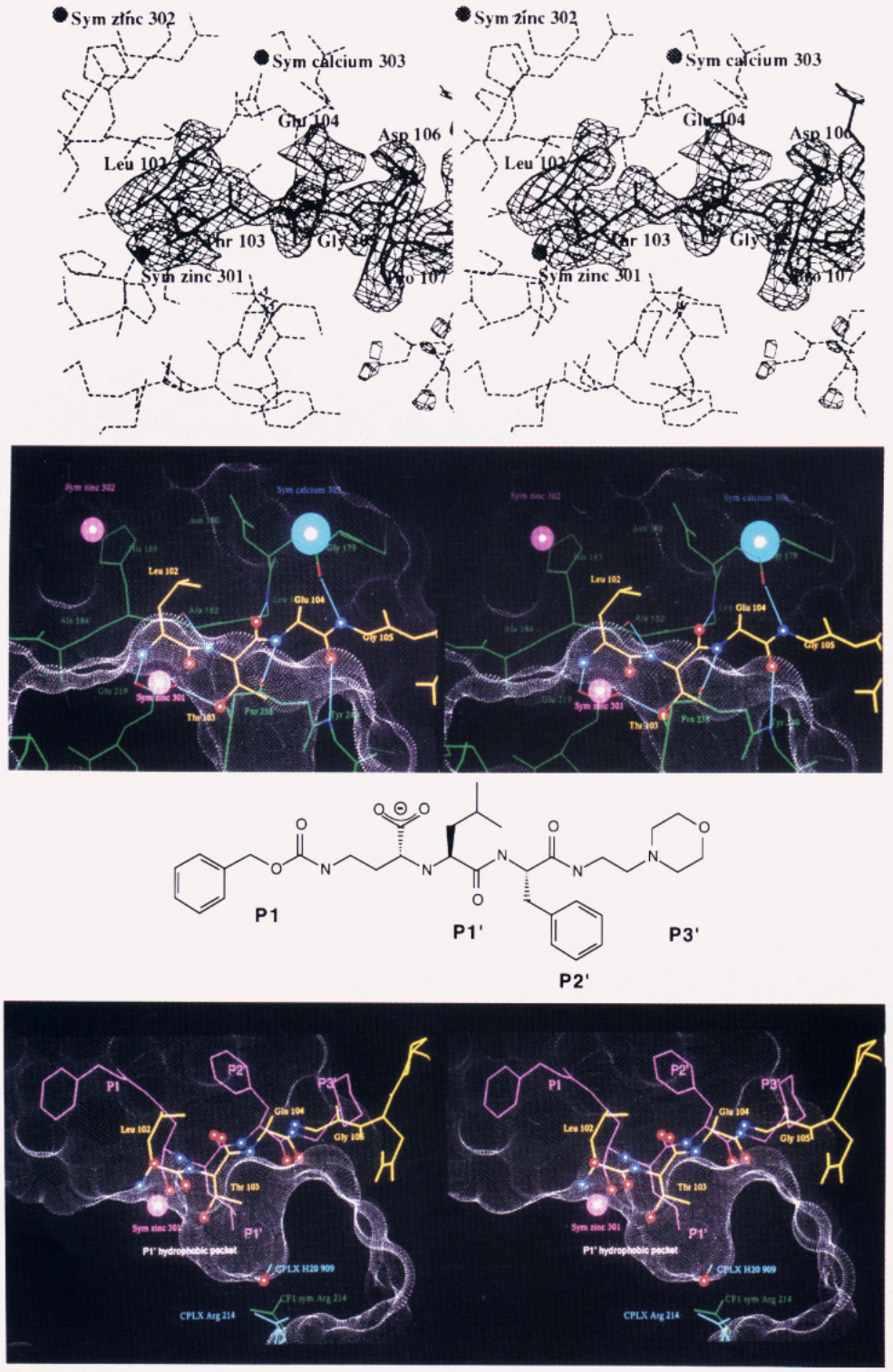

FIGURE 3: (A, top) An $F_o - F_c$ simulated-annealing omit map (Hodel et al., 1992) of the amino terminus of CF2 molecule A contoured at 2.5σ. For this map, residues 102–107 of CF2 molecule A were omitted, and the remaining portions of CF2 were refined using the SA slow-cooling protocol (Brünger et al., 1990). Residues 102–107 are drawn with bold solid lines while the crystallographic symmetry-related active site cleft is represented by dashed lines. (B, next to top) Interactions between the CF1 amino terminus (yellow) and a crystallographic symmetry-related molecule (green). The symmetry-related collagenase active site cleft is mapped with a white Connally dot surface (with a rolling sphere of 1.4 Å-radius). Symmetry-related residues are labeled in green, and CF1 N-terminal residues are labeled in yellow. CF1 atoms that either ligate *sym* zinc 301 or form intermolecular hydrogen bonds are shown as red (oxygen) or blue (nitrogen) spheres. The seven intermolecular hydrogen bonds that are present in CF1 and CF2 (Table 4) are represented by blue lines. (C, next to bottom) Structure of the CPLX inhibitor. P1′ is a leucine, P2′ is a phenylalanine, P3′ is a morpholino group, and P1 is a benzyloxycarbonylamino group. (D, bottom) The CF1 amino terminus in the active site of a crystallographic symmetry-related molecule. The CPLX inhibitor (Figure 3C) is superimposed onto the CF1 amino terminus using the Cα positions of CPLX molecule A and the CF1 crystallographic symmetry mate. The LSQ_IMPROVE option of the program O (Jones, 1990) was used for this superposition. CF1 atoms that participate in intermolecular hydrogen bonds or ligate the symmetry-related catalytic zinc (*sym* zinc 301) are drawn as red (oxygen) or blue (nitrogen) spheres. Corresponding atoms in the CPLX inhibitor are drawn as spheres with the same coloring scheme. In all collagenase crystal forms, the guanidinium group of Arg 214 forms the base of the P1′ pocket. In CPLX, CF1, and CF2, a water molecule (CPLX H_2O 909) is present in the P1′ pocket, interacting with Arg 214.

proteins contain a catalytic domain that shares structural resemblance with 19-kDa collagenase. The profile search did not assign high scores to thermolysin or asctacin (Z scores < 1.0).

Table 4: Intermolecular Hydrogen Bonds Found in the Active Site Cleft of CF1, CF2, and CPLX[a]

		CF1 and CF2[b]				CPLX		
	active site cleft			CF2 (Å)			CPLX (Å)	
type of interaction	atom	atom	CF1	mol A	mol B	atom	mol A	mol B
hydrogen bonds	Glu 219 Oε2	Leu 102 N	3.0	3.0	2.9			
	Ala 182 CO	Thr 103 N	3.3	3.2	3.2	P1′ N	3.0	2.8
	Glu 219 Oε1	Thr 103 Oγ1	2.6	2.7	3.1			
	Leu 181 N	Thr 103 CO	2.9	2.9	2.8	P1′ CO	2.8	2.8
	Pro 238 CO	Glu 104 N	2.8	3.2	3.1	P2′ N	3.8[c]	3.7[c]
	Tyr 240 N	Glu 104 CO	2.9	3.0	3.3	P2′ CO	3.0	2.7
	Gly 179 CO	Gly 105 N	2.8	2.7	2.7	P3′ N	3.1	3.0
	Glu 219 Oε1					P1′ N	3.2	3.4
	Asn 180 Oδ1					P1 N	3.0	2.9
	Asn 180 Nδ2					P1 CO	2.7	2.9
	His 183 Nδ1					P1 CO	2.7	2.7
zinc ligating	zinc 301	Leu 102 N	1.8	2.1	2.1			
		Leu 102 CO	2.1	2.1	2.2			
						carboxy Oε1	2.0	2.1
						carboxy Oε2	2.6	2.7

[a] Each row lists a bond that occurs in the active site cleft of CF1, CF2, or CPLX. Bonds that are comparable in CF1, CF2, and CPLX are on the same row. Bonds that are not comparable are listed individually. [b] In CF1 and CF2, the amino terminus forms a complex with the active site cleft of a symmetry-related molecule. [c] In CPLX, P2′ N and Pro 238 CO do not interact with each other.

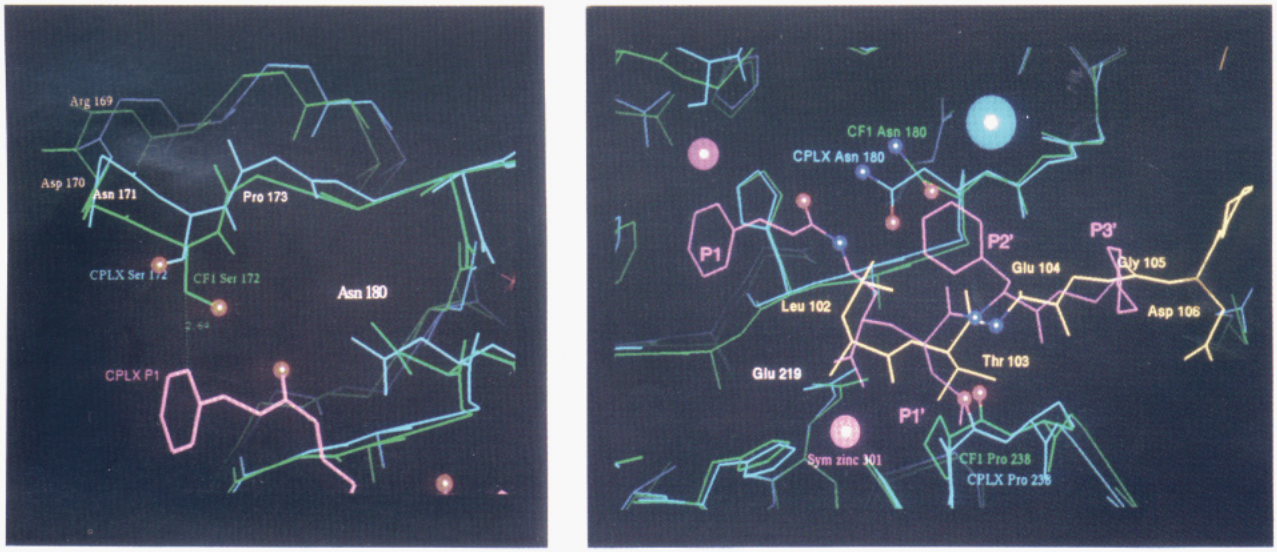

FIGURE 4: Comparison of the CF1 and CPLX active site clefts. (A, left) Superposition of CPLX (blue) onto a CF1 crystallographic symmetry mate (green) using the matrix described in the caption of Figure 3D. The CPLX inhibitor (Figure 3C) appears to "push up" the loop between β strands III and IV (residues 170–173). The steric clash that would arise between the CPLX inhibitor and the CF1 active site cleft is highlighted by the distance between Cβ of CF1 Ser 172 and Cδ2 of the CPLX P1 phenyl ring (2.6 Å), which is represented by a dotted green line. The P1 phenyl ring prevents the formation of a hydrogen bond between the Ser 172 hydroxy and P1 carbonyl groups (red spheres). (B, right) The amino terminus of CF1 (yellow) in the active site cleft of a symmetry-related molecule (green). CPLX molecule A (blue) is superimposed onto the CF1 crystallographic symmetry mate (green) using the matrix described in the caption of Figure 3D. The χ_1 angle of Asn 180 is gauche$^+$ in CF1 and CF2 and trans in CPLX. A Leu occupies the P1′ pocket in CPLX, and a Thr occupies the P1′ pocket in CF1 and CF2. In CPLX, the P1′ Leu appears to push Pro 238 outward, and a potential interaction between Pro 238 CO (red spheres) and P2′ N (blue sphere) is lost. A hydrogen bond forms between the corresponding atoms in CF1 and CF2 (Table 4).

Comparison of the Thermolysin and Collagenase Reaction Mechanisms. The structure of thermolysin complexed to a series of inhibitors has led to the proposed reaction mechanism for the thermolysin-catalyzed cleavage of peptides (Matthews, 1988; Figure 5A). In the thermolysin mechanism, Glu 143TL acts as a general base that, together with the catalytic zinc, coordinates a water molecule near the carbonyl of the scissile bond. The coordinated water molecule attacks the carbonyl carbon of the scissile bond and donates a proton to Glu 143TL. Tyr 157 OH, the positively charged catalytic zinc, and His 231TL Nε2 stabilize the resultant negative charge that forms at the scissile carbon. Glu 143TL then transfers the donated proton to the scissile nitrogen while the carbonyl of Ala 113TL and Asn 112TL Oδ1 stabilize the resulting positive charge on the scissile nitrogen. Glu 143TL then shuttles the remaining proton from the water molecule to the scissile nitrogen, resulting in cleavage of the peptide bond (Matthews, 1988).

The structure of 19-kDa collagenase suggests that some functional groups in the thermolysin reaction mechanism may be conserved in collagenase (Figure 5B). Collagenase Glu 219, which corresponds to Glu 143TL (Figure 5), forms a hydrogen bond with the scissile nitrogen analog of the inhibitor, in CPLX, and the N-terminus, in CF1 and CF2 (Table 4). Collagenase Glu 219 represents the glutamate of the HEXGH motif that is conserved in MMPs (Birkedal-Hansen et al., 1993; Table 6). The interaction of Glu 219 with atoms that mimic the main-chain nitrogen atoms of substrates in CF1, CF2, and CPLX suggests that Glu 219 plays a role comparable to Glu 143TL. However, the presence of Leu 103 and Thr 104 in the catalytic cleft of a symmetry-related molecule in the CF1 and CF2 crystal forms raises the question as to why the Leu 103–Thr 104 peptide bond was not cleaved. The lack

Table 5: Results of a Profile Search Using the Method of Bowie et al. (1991) and a Probe Consisting of Residues 107–263 of CF1[a]

protein	% identity	*Z* score
collagenase, porcine	89	34.9
collagenase, bovine	87	33.9
fibroblast collagenase, human (MMP-1)	100	33.5
collagenase, rabbit	88	31.8
neutrophil collagenase, human (MMP-8)	67	28.4
collagenase, rat	56	25.7
transformation-associated protein, rat	58	23.1
stromelysin-1, mouse	62	23.0
uterine metalloprotease, human (MMP-7, PUMP)	52	23.0
macrophage metalloelastase, mouse	54	22.2
stromelysin-1, rabbit	62	22.0
stromelysin-1, rat	59	21.2
stromelysin-1, human (MMP-3)	59	21.0
stromelysin-2, human (MMP-10)	58	20.9
gelatinase, 92 kDa, human (MMP-9)	52	18.5
gelatinase, 72 kDa, human (MMP-2)	52	18.1
gelatinase, 77 kDa, mouse	52	18.0
stromelysin-3, human (MMP-11)	47	14.5
soybean metalloendoproteinase (SMEP)	45	13.7
hatching enzyme precursor (suHEP)	43	11.8

[a] Percent identity is the percentage of identical amino acids in the sequence of the target protein with residues 101–270 of human fibroblast collagenase after alignment with the program BESTFIT (Devereux, 1984). The *Z* score for each sequence is the number of standard deviations above the mean alignment score for other sequences of similar length (Gribskov et al., 1990). All proteins with *Z* scores greater than 8.6 are listed. For the search, a gap-opening penalty of 4.5 and a gap-extension penalty of 0.05 was used.

of cleavage between Leu 103 and Thr 104 is particularly interesting since analysis of collagenase samples used in CF1 and CF2 crystallization trials, by mass spectrometry and N-terminal sequencing, indicated that as many as four residues (Met 100 through Thr 103) are missing (Hassell et al., 1994). In CF1 and CF2, the structure clearly shows that the amino terminus is bound in an inhibitory binding mode. By having Thr 103 Oγ hydrogen bonding to the catalytically important Glu 219, hydrolysis is disrupted. N-Terminal analysis of freshly prepared 19-kDa collagenase shows it to be a mixture of full length, N-1, and N-2. Only after several weeks are significant amounts of further degradation products, corresponding to the Leu 102–Thr 103 bond, seen. In thermolysin, evidence for the catalytic water molecule (Figure 5) includes the presence of a water molecule that interacts with the main-chain nitrogen of Trp 115TL in several thermolysin–inhibitor complexes (Matthews, 1988). In all five molecules of collagenase, a comparable water molecule is interacting with the main-chain nitrogen of the residue that corresponds to Trp 115TL, Ala 184 (Figure 6).

The carbonyl of collagenase Ala 182 (Figure 3B), which corresponds to Ala 113TL (Figure 5), forms hydrogen bonds with the scissile nitrogen analog, in CPLX, and with the main-chain nitrogen of Thr 103, in CF1 and CF2 (Table 4). If Ala 182 plays a role in the collagenase-catalyzed reaction mechanism, three lines of evidence suggest that the precise position of the carbonyl of Ala 182 is important. First, residues 181–184 are conserved in MMPs and, after superposition, are indistinguishable in CF1, CF2, and CPLX (Table 6). Conservation of Ala 182 may be important because the residue is in a β strand (IV) with the Cβ carbon completely buried. As a result, substitution of Ala 182 for a larger residue would cause β strand IV to bulge and disrupt the carbonyl of Ala 182. Second, although collagenase residues 182–184 and thermolysin residues 113TL to 115TL are superimposable, the peptide chain before and after this segment is not. Similarly, Gomis-Rüth et al. (1993) found that the segment corresponding to collagenase residues 182–184 in thermolysin and astacin (64AS to 66AS, 113TL to 115TL) shares structural agreement while the segments before and after this region do not. Finally, the portion of collagenase that lies directly under Ala 182 (residues 196–198) is conserved in MMPs (Table 6) and, after superposition, is indistinguishable in the collagenase structures. The conservation of the region below Ala 182 may be important because a Phe197Ala mutation, for example, would cause Ala 182 to collapse into the hole created by the absence of the Phe 197 phenyl ring and disrupt the position of Ala 182 CO.

Not all the groups that are involved in the reaction mechanism proposed by Matthews (1988) have functional counterparts in collagenase. Collagenase Asn 180, which corresponds to Asn 112TL (Figure 5), does not interact significantly with the inhibitor. The distance between Oδ1 of collagenase Asn 180 and the scissile nitrogen analog is 3.9 Å in CPLX (P1′N) and 6.1 Å in CF1 (Thr 103 N). Favorable interactions between Asn 180 Oδ1 and the scissile nitrogen are also precluded by the proximity of P1′ CO in CPLX (3.2 Å). These observations indicate that Asn 180 does not have a significant role in the collagenase-catalyzed reaction mechanism. Because Asn 180 is not conserved in MMPs, it cannot be included in a general MMP mechanism but may provide a source of MMP substrate specificity. Another thermolysin residue implicated in the thermolysin reaction mechanism, Tyr 157TL (Figure 5), is in a loop that extends from the N-terminal side of the active site helix. This loop is bordered by the active site ligands His 146TL and Glu 166TL. The corresponding loop in collagenase, between His 222 and His 228, does not contain any group that can substitute for Tyr 157TL. Also, there is no group analogous to Tyr 157TL in astacin (Bode et al., 1992).

There is no residue in 19-kDa collagenase that is comparable to His 231TL (Figure 5). Astacin Tyr 149AS appears to play a role that is comparable to His 231TL (Bode et al., 1992). His 231TL and Tyr 149AS are in the C-terminal domain. Since 19-kDa collagenase has a truncated C-terminal domain, it cannot be used to identify residues in the C-terminal domain that are equivalent to His 231TL or Tyr 149AS. Some observations suggest that there are no such residues in full-length collagenase. First, there are two naturally occurring metalloproteinases that have no C-terminal domain. These are soybean metalloendoproteinase, SMEP (McGeehan et al., 1992), and the mammalian enzyme, uterine metalloprotease (PUMP; Muller et al., 1988). Second, truncated forms of MMPs have been expressed and characterized that have the same catalytic efficiency as full-length collagenase toward peptide substrates (Murphy et al., 1992; D. Becherer, A. S. Howe, A. Srinivasan, I. Patel, G. B. Wisely, H. LeVine, and G. McGeehan, personal communication). These observations suggest that C-terminal-based residues do not play a crucial role in the enzymatic activity of collagenase. However, C-terminal-based residues could play a role in generating the substrate specificity that is exhibited by MMPs.

Since MMP catalytic zincs are ligated by three neutral histidines rather than the two histidines and a negatively charged glutamate found in thermolysin (Matthews et al., 1972), the MMP catalytic zincs are predicted to have a higher net positive charge and a greater ability to stabilize the negative charge present in the transition state than the thermolysin catalytic zinc. It is possible that MMPs have evolved to have a more catalytically efficient zinc rather than a dependence on the C-terminal domain based charge-stabilizing groups found in thermolysin and astacin.

Table 6: Secondary Structure of 19-kDa Collagenase and Sequence Alignment with Other MMPs[a]

		I	A	II
Structure				
MMP-1	102	LTEGNPRWEQ	THLRYRIENY TPDLPRADVD	HAIEKAFQLW SNVTPLTFTK
MMP-2	112	FFPRKPKWDK	NQITYRIIGY TPDLDPETVD	DAFARAFQVW SDVTPLRFSR
MMP-3	105	...GIPKWRK	THLTYRIVNY TPDLPKDAVD	SAVEKALKVW EEVTPLTFSR
MMP-7	97	LFPNSPKWTS	KVVTYRIVSY TRDLPHITVD	RLVSKALNMW GKEIPLHFRK
MMP-8	101	LTPGNPKWER	TNLTYRIRNY TPQLSEAEVE	RAIKDAFELW SVASPLIFTR
MMP-9	111	..EGDLKWHH	HNITYWIQNY SEDLPRAVID	DAFARAFALW SAVTPLTFTR
MMP-10	104	...GMPKWRK	THLTYRIVNY TPDLPRDAVD	SAIEKALKVW EEVTPLTFSR
MMP-11	98	FVLSGGRWEK	TDLTYRILRF PWQLVQEQVR	QTMAEALKVW SDVTPLTFTE
suHEP	169	TSSITWSRN	QPVTYSFGAL TSDLNQNDVK	DEIRRAFRVW DDVSGLSFRE

		III	IV	V
Structure				
MMP-1	152	VSEGQADIMI	SFVRGD.HRD NSPFDGPGGN	LAHAFQPGPG IGGDAHFDEH
MMP-2	162	IHDGEADIMI	NFGRWE.HGD GYPFDGKDGL	LAHAFAPGTG VGGDSHFDDD
MMP-3	152	LYEGEADIMI	SFAVRE.HGD FYPFDGPGNV	LAHAYAPGPG INGDAHFDDD
MMP-7	147	VVWGTADIMI	GFARGA.HGD SYPFDGPGNT	LAHAFAPGTG LGGDAHFDED
MMP-8	151	ISQGEADINI	AFYQRD.HGD NSPFDGPNGI	LAHAFQPGQG IGGDAHFDAE
MMP-9	159	VYSRDADIVI	QFGVAE.HGD GYPFDGKDGL	LAHAFPPGPG IQGDAHFDDD
MMP-10	151	LYEGEADIMI	SFAVKE.HGD FYSFDGPGHS	LAHAYPPGPG LYGDIHFDDD
MMP-11	148	VHEGRADIMI	DFARYW.DGD DLPFDGPGGI	LAHAFFPKTH REGDVHFDYD
suHEP	218	VPDTTSVDIR	IKFGSYDHGD GISFDGRGGV	LAHAFLP... RNGDAHFDDS

		B	
Structure			
MMP-1	201	ERWTNNFTE.	YNLHRVAAHE LGHSLGLSHS TDIGALMYPS Y..TF..SGD
MMP-2	211	ELWTLGEGQ*	YSLFLVAAHE FGHAMGLEHS QDPGALMAPI Y..TY..TKN
MMP-3	201	EQWTKDTTG.	TNLFLVAAHE IGHSLGLFHS ANTEALMYPL Y..HS..LTD
MMP-7	196	ERWTDGSSLG	INFLYAATHE LGHSLGMGHS SDPNAVMYPT Y..GNGDPQN
MMP-8	200	ETWTNTSAN.	YNLFLVAAHE FGHSLGLAHS SDPGALMYPN Y..AFRETSN
MMP-9	208	ELWSLGKG*G	YSLFLVAAHE FGHALGLDHS SVPEALMYPM Y..RF..TEG
MMP-10	200	EKWTEDAS.G	TNLFLVAAHE LGHSLGLFHS ANTEALMYPL YNSFT...EL
MMP-11	197	ETWTIGDDQG	TDLLQVAAHE FGHVLGLQHT TAAKALMSAF Y..TF..RYP
suHEP	265	ETWTEGTRSG	TNLFQVAAHE FGHSLGLYHS TVRSALMYPY Y.QGY..VPN

		C		
Structure				
MMP-1	246	V...QLAQDD	IDGIQAIYGR	SQNPVQPI 270
MMP-2	431	F...RLSQDD	IKGIQELYGA	SPDIDLGT 455
MMP-3	246	LTRFRLSQDD	INGIQSLYGP	PPDSP 270
MMP-7	244	F.. KLSQDD	IKGIQKLYGK	RSNSRKK 267
MMP-8	247	Y...SLPQDD	IDGIQAIYGL	SSNPIQP 270
MMP-9	429	P...PLHKDD	VNGIRHLYGP	RPEPEPR 452
MMP-10	246	A.QFRLSQDD	VNGIQSLYGP	PPASTEE 271
MMP-11	243	L SLSPDD	CRGVQHLYGQ	PWPTVTS 266
suHEP	312	F...RLDNDD	IAGIRSLYGS	NSGSGTTT 336

[a] In 19-kDa fibroblast collagenase (MMP-1) β strand I spans residues 113–118, α helix A spans residues 129–144, β strand II spans residues 147–153, β strand III spans residues 158–164, β strand IV spans residues 182–185, β strand V spans residues 194–199, α helix B (the active site helix) spans residues 212–223, and α helix C spans residues 250–260. β strands and α helices are represented by blue and red highlighting, respectively. The fibronectin-like inserts (MMP-2 residues 220–394 and MMP-9 residues 216–391) are represented by yellow stars. MMP residues that correspond to MMP-1 catalytic zinc ligands (His 218, His 222, and His 228) are in blue. MMP residues that correspond to the MMP-1 secondary zinc ligands (His 168, Asp 170, His 183, and His 196) as well as the residues that ligate calcium 303 through side-chain interactions (Asp 175, Asp 198, and Glu 201) are in red. MMP residues that correspond to MMP-1 Glu 219 are in yellow. Residues that are in close proximity of MMP-1 Ala 182 and are structurally conserved in CF1, CF2, and CPLX (residues 181–184 and residues 196–198) are highlighted in green.

CONCLUSIONS

The aberrant regulation of MMPs in pathologies such as arthritis and metastasis has prompted a tremendous effort to develop drugs capable of inhibiting MMP activity (Henderson et al., 1990). One common approach to designing MMP inhibitors is the use of a zinc binding group, such as a thiol or carboxylate, based in a peptide framework modeled after the sequences of MMP cleavage sites. Prior to the elucidation of a MMP structure (Lovejoy et al., 1994), structure-based approaches have focused largely on modeling the active site of MMPs on the basis of the structure of thermolysin (Matthews et al., 1972). The structure of 19-kDa collagenase forming two different complexes provides new insight on the flexibility of the active site cleft. This knowledge can be used to design collagenase inhibitors that have greater binding affinity. Comparison of the complex-stabilizing interactions that are conserved in CF1, CF2, and CPLX also provides insight on the groups that may be involved in the collagenase-catalyzed cleavage of substrates.

The molecular biology of MMPs is complex, and there is little direct evidence to support the involvement of particular MMPs in specific biological processes (Birkedal-Hansen et al., 1993). MMPs are regulated by several growth factors and hormones, as well as the tissue inhibitor of the metalloproteinase (TIMP) family of proteins (Birkedal-Hansen et al., 1993; Woessner, 1991). Individual members of the MMP family have broad substrate specificity. The crystal structure of collagenase will facilitate the detailed modeling of other MMPs. These models will provide new direction for the design of compounds that selectively inhibit individual members of

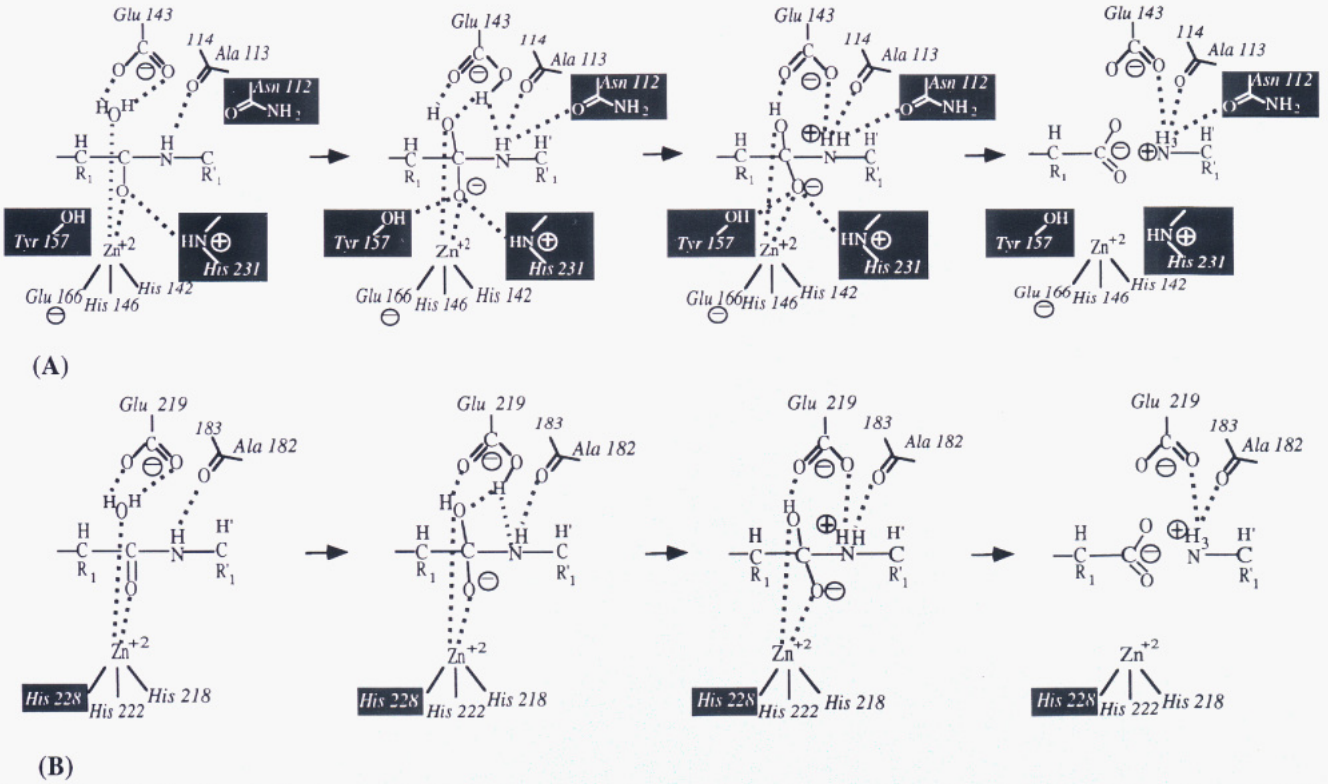

FIGURE 5: (A) Reaction mechanism for the thermolysin-catalyzed cleavage of peptides (Matthews, 1988). (B) Groups in collagenase that may have similar properties as their thermolysin counterparts. Functional groups that are unique to either thermolysin or collagenase are highlighted.

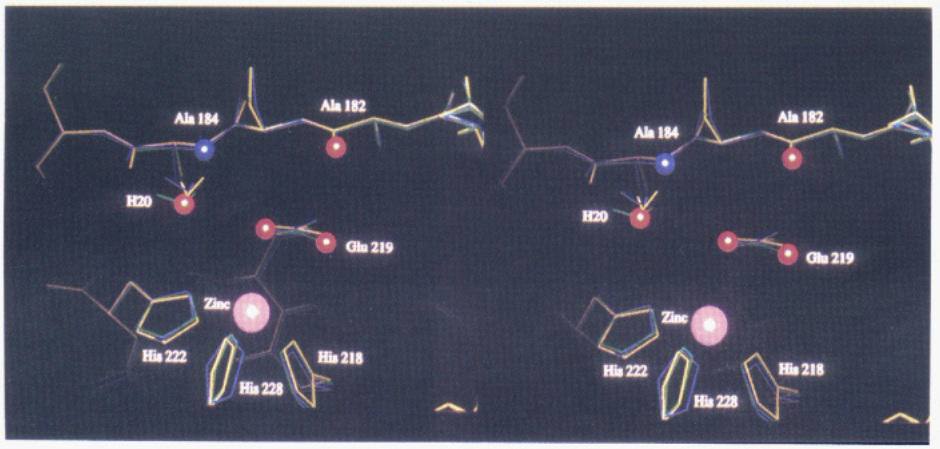

FIGURE 6: A view of the collagenase groups that may have properties similar to their thermolysin counterparts. In CF1 (green) and CF2 molecules A (blue) and B (yellow) as well as CPLX (not shown) a water molecule is associated with the main-chain nitrogen of Ala 184. It is possible that this water molecule, when coordinated by Glu 219, acts as a nucleophile and attacks the carbonyl carbon of the scissile bond. CF1 and CF2 were superimposed as described in Table 3.

the MMP family. In addition to potentially serving as treatments for arthritis and related diseases, such inhibitors will help determine the exact role each MMP serves in a wide range of physiological processes.

ACKNOWLEDGMENT

We thank K. Y. J. Zhang and D. Eisenberg for helpful comments and for the use of their three-dimensional profile method. We thank M. Milburn for determining the space group of CF1 and initial CF1 data collection. We also thank M. Lambert and G. McGeehan for helpful discussions. Color figures were made with Insight II (Biosym Technologies, Inc.). We acknowledge the recent published reports on the structure of the catalytic domains of human fibroblast collagenase (Borkakoti et al., 1994), human neutrophil collagenase (Strams et al., 1994; Bode et al., 1994), stromelysin-1 (Gooley et al., 1994), and matrilysin (PUMP) (Browner, 1994).

REFERENCES

Becherer, J. D., Howe, A. S., Patel, I., Wisely, B., Levine, H., & McGeehan, G. M. (1991) *J. Cell. Biochem., Suppl. 15G*, 139.

Berman, J., Green, M., Sugg, E., Anderegg, R., Millington, D. S., Norwood, D. L., McGeehan, J., & Wiseman, J. (1992) *J. Biol. Chem. 267*, 1434.

Birkedal-Hansen, H. (1988) *J. Oral Pathol. 17*, 445.

Birkekal-Hansen, H., Moore, W. G. I., Bodden, M. K., Windsor, L. J., Birkedal-Hansen, B., DeCarlo, A., & Engler, J. A. (1993) *Crit. Rev. Oral Biol. Med. 4*, 197.

Bode, W., Gomis-Rüth, F. X., Huber, R., Zwilling, R., & Stöcker, W. (1992) *Nature 358*, 164.

Bode, W., Reinemer, P., Huber, R., Kleine, T., Schnierer, S., & Tschesche, H. *EMBO J.* (in press).

Borkakoti, N., Winkler, F. K., Williams, D. H., D'Arcy, A., Broadhurst, M. J., Brown, P. A., Johnson, W. H., & Murray, E. J. (1994) *Struct. Biol. 1*, 106.

Bowie, J. U., Lüthy, R., & Eisenberg, D. (1991) *Science 253*, 164.
Browner, M. (1994) *J. Cell. Biochem., Suppl. 18D*, 120.
Brünger, A. T. (1990) *Acta Crystallogr. A46*, 46.
Brünger, A. T., Krukowski, A., & Erikson, J. W. (1990) *Acta Crystallogr. A46*, 585.
Cawston, T. E., Mercer, E., de-Silvia, E., & Hazleman, B. L. (1984) *Arthritis Rheum. 27*, 285.
Collier, I. E., Krasnov, P. A., Strongin, A. Y., Birkedal-Hansen, H., & Goldberg, G. I. (1992) *J. Biol. Chem. 267*, 6776.
Devereux, J., Haeberli, P., & Smithies, O. (1984) *Nucleic Acids Res. 12*, 387.
Gomis-Rüth, F. X., Stöcker, W., Huber, R., Zwilling, R., & Bode, W. (1993) *J. Mol. Biol. 229*, 945.
Gooley, P. R., O'Connell, J. 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. (1994) *Struct. Biol. 1*, 111.
Gribskov, M., Lüthy, R., & Eisenberg, D. (1990) *Methods Enzymol. 183*, 146.
Hassell, A. M., Anderegg, R. J., Weigl, D., Milburn, M. V., Burkhart, W., Luther, M. A., & Jordan, S. R. (1994) *J. Mol. Biol. 236*, 1410.
Hasty, K. A., Jeffrey, J. J., Hibbs, M. S., & Welgus, H. G. (1987) *J. Biol. Chem. 262*, 10048.
Hayakawa, T., Yamashita, K., Kodama, S., Iwata, H., & Iwata, K. (1991) *Biomed. Res. 12*, 169.
Henderson, B., Docherty, A. J. P., & Beeley, N. R. A. (1990) *Drugs Future 15*, 495.
Hendrickson, W. A. (1979) *Acta Crystallogr. A35*, 158.
Hodel, A., Kim, S.-H., & Brünger, A. T. (1992) *Acta Crystallogr. A48*, 841.
Jones, T. A. (1985) *Methods Enzymol. 115*, 157.
Jones, T. A. (1990) *O, Version 5, Users Manual*, Department of Molecular Biology, BMC, Box 590, S-75124, Uppsala, Sweden.
LePage, T., & Gache, C. (1990) *EMBO J. 9*, 3003.
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. (1994) *Science 263*, 375.
Lowry, C. L., McGeehan, G., & LeVine, H., III (1992) *Proteins: Struct., Funct., Genet. 12*, 42.
Matthews, B. W. (1988) *Acc. Chem. Res. 21*, 333.
Matthews, B. W., Jansonius, J. N., Colman, P. M., Schoenborn, B. P., & Dupourge, D. (1972) *Nature, New Biol. 238*, 37.
McGeehan, G., Burkhart, W., Anderegg, R., Becherer, J. D., Gillikin, J. W., & Graham, J. S. (1992) *Plant Physiol. 99*, 1179.
Monzingo, A. F., & Matthews, B. W. (1984) *Biochemistry 23*, 5724.
Muller, D., Quantin, B., Gesnel, M. C., MillonCollard, R., Abecassis, J., & Breathnach, R. (1988) *Biochem. J. 253*, 187.
Murphy, G., Allan, J. A., Willenbrock, F., Cockett, M. I., O'Connell, J. P., & Docherty, A. J. P. (1992) *J. Biol. Chem. 267*, 9612.
Nagase, H., Barrett, A. J., & Woessner, J. F., Jr. (1992) *Matrix, Suppl. 1*, 421.
Rao, S. N., Jih, J.-H., & Hartsuck, J. A. (1980) *Acta Crystallogr. A36*, 878.
Stams, T., Spurlino, J. C., Smith, D. L., Wahl, R. C., Ho, T. F., Qoronfleh, M. W., Banks, T. M., & Rubin, B. (1994) *Struct. Biol. 1*, 119.
Welgus, H. G., Jeffrey, J. J., & Eisen, A. Z. (1981) *J. Biol. Chem. 256*, 9511.
Windsor, L. J., Birkedal-Hansen, H., Birkedal-Hansen, B., & Engler, J. A. (1991) *Biochemistry 30*, 641.
Woessner, J. F., Jr. (1991) *FASEB J. 5*, 2145.
Woolley, D. E., Harris, E. D., Jr., Mainardi, C. L., & Brinckerhoff, C. E. (1978) *Science 200*, 773.
Zucker, S., Lysik, R. M., Zarrabi, M. H., & Moll, U. (1993) *Cancer Res. 53*, 140.