Crystal Structure of an Ecotin−Collagenase Complex Suggests a Model for Recognition and Cleavage of the Collagen Triple Helix†

John J. Perona,‡,§ Christopher A. Tsu,‡,∥ Charles S. Craik,*,‡,⊥ and Robert J. Fletterick⊥

Departments of Pharmaceutical Chemistry and Biochemistry and Biophysics, University of California, San Francisco, California 94143-0446

Received July 17, 1996; Revised Manuscript Received December 10, 1996⊗

ABSTRACT: The crystal structure of fiddler crab collagenase complexed with the dimeric serine protease inhibitor ecotin at 2.5 Å resolution reveals an extended cleft providing binding sites for at least 11 contiguous substrate residues. Comparison of the positions of nine intermolecular main chain hydrogen bonding interactions in the cleft, with the known sequences at the cleavage site of type I collagen, suggests that the protease binding loop of ecotin adopts a conformation mimicking that of the cleaved strand of collagen. A well-defined groove extending across the binding surface of the enzyme readily accommodates the two other polypeptide chains of the triple-helical substrate. These observations permit construction of a detailed molecular model for collagen recognition and cleavage by this invertebrate serine protease. Ecotin undergoes a pronounced internal structural rearrangement which permits binding in the observed conformation. The capacity for such rearrangement appears to be a key determinant of its ability to inhibit a wide range of serine proteases.

Collagen is an essential component of the extracellular matrix in higher organisms, and its degradation by specific proteases is a key step in connective tissue remodeling (Woessner, 1991). In vertebrate species, the family of enzymes known as the matrix metalloproteases are responsible for collagen cleavage. These enzymes possess an N-terminal 80-amino acid propeptide which is cleaved to generate mature enzyme, a catalytic domain of some 180 amino acids, and a C-terminal hemopexin-like domain providing collagen binding determinants (Hirose et al., 1993; Sanchez-Lopez et al., 1993). Crystal structures of the catalytic domains of human fibroblast collagenase (Borkakoti et al., 1994; Lovejoy et al., 1994; Spurlino et al., 1994) and human neutrophil collagenase (Stams et al., 1994; Bode et al., 1994) have revealed a strong similarity to the bacterial metalloprotease thermolysin (Matthews et al., 1972). Two structures of the hemopexin domain, both alone (Faber et al., 1995) and as part of the full-length porcine synovial collagenase (Li et al., 1995), have also recently appeared. However, the structural determinants conferring collagenolytic specificity, and the means by which recognition of the triple helix is coupled to cleavage, remain speculative. Collagen possesses a tightly wound fibrous structure of three extended, left-handed polyproline II-like helices (Rich & Crick, 1961; Fraser et al., 1979; Bornstein & Traub, 1979; Bella et al., 1994). Some melting of this triple helix would appear to be a likely requirement for cleavage to occur.

Certain members of the chymotrypsin-like serine protease family are also capable of cleaving triple-helical type I collagen. Enzymes of invertebrate origin, for example, the *Hypoderma lineatum* collagenase (Lecroisey et al., 1987), shrimp chymotrypsin (Sellos & Van Wormhaut, 1992), and *Uca pugilator* (fiddler crab) collagenase (Grant et al., 1980; Tsu et al., 1994), each cleave bonds within a subdomain of type I collagen located some three-fourths of the distance from the N termini of the roughly 1000 amino acid chains. The cleavage sites of fiddler crab collagenase in type I soluble calf skin collagen have been mapped in detail and found to be directly adjacent to those of the mammalian extracellular matrix enzymes (Tsu et al., 1994). Elucidation of collagen specificity determinants within enzymes of the well-characterized serine protease family thus provides an attractive complement to the study of the vertebrate metalloenzymes. Of further interest is the comparative structural analysis among members of the chymotrypsin-like proteases which either can or cannot cleave type I collagen. This comparison poses interesting questions in protein−protein recognition and in the molecular evolution of the serine proteases.

Ecotin is a general inhibitor of pancreatic serine proteases found in the bacterial periplasm and is classified as a member of the small substrate-like inhibitors exemplified by bovine pancreatic trypsin inhibitor (BPTI) (Chung et al., 1983; McGrath et al., 1991). The molecule exists as a dimeric species; each 142-amino acid monomer targets one protease molecule. An exceptionally broad range of serine proteases are effectively inhibited, including fiddler crab collagenase as well as each of the mammalian enzymes trypsin, chymotrypsin, elastase, and factor Xa. K_i values for the interactions are each in the range of 1 nM to 50 pM (Chung et al., 1983; Seymour et al., 1994; Pal et al., 1994). The crystal structure of ecotin complexed with rat trypsin reveals that the molecule possesses a modified jellyroll β-sheet fold, with a 20-amino acid C-terminal extension which enclasps a second monomer to form head-to-tail dimers (McGrath et al., 1994). The

† Supported by NIH Postdoctoral Fellowship GM13818-03 (to J.J.P.), NIH Grant DK39304 (to R.J.F.), and NSF Grant MCB-9604379 (to C.S.C.).

* Author to whom correspondence should be addressed. Phone: 415-476-8146. Fax: 415-476-0688.

‡ Department of Pharmaceutical Chemistry.

§ Present address: Department of Chemistry and Interdepartmental Program in Biochemistry and Molecular Biology, University of California, Santa Barbara, CA 93106-9510.

∥ Present address: Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309.

⊥ Department of Biochemistry and Biophysics.

⊗ Abstract published in *Advance ACS Abstracts,* April 1, 1997.

S0006-2960(96)01752-7 CCC: $14.00

ecotin dimer interacts with two trypsin molecules to form a tetrameric complex which features a second discrete binding site on the enzyme. This results in effective chelation of the protease with a consequent augmentation of binding energy. The existence of the secondary site provides one rationalization for the pan specificity of the inhibitor, although the mechanism by which generalized binding occurs has remained obscure.

To further explore the structural basis for the collagenolytic specificity and broad inhibitory profile of crab collagenase and ecotin, respectively, we have determined the crystal structure of this protein−protein complex at 2.5 Å resolution. We find that when bound to collagenase the structure of the ecotin monomer is strikingly rearranged relative to its conformation bound to trypsin. Several β-strands of ecotin are displaced by up to 6 Å to allow protease interactions which extend from the S7 to the S4′ sites. The internal flexibility of ecotin appears to be a key structural determinant underlying its exceptionally broad and potent inhibitory profile. On the basis of a correlation with collagen cleavage site mapping, the primary binding loop of ecotin is suggested to adopt a conformation similar to that of the cleaved strand of collagen. This correlation provided the initial impetus for construction of a detailed model of the interaction of triple-helical collagen with fiddler crab collagenase.

EXPERIMENTAL PROCEDURES

Protein Purification and Determination of the Inhibition Constant. Ecotin and crab collagenase were purified as described previously (McGrath et al., 1991; Tsu et al., 1994). The ecotin and collagenase were incubated for 10 min in a buffer solution containing 50 mM Tris (pH 8.0), 100 mM NaCl, and 20 mM $CaCl_2$ at 25 °C. The enzyme concentration was 122 pM, and the concentration of the inhibitor was varied between 0 and 2000 pM. Inhibition was measured spectrophotometrically at 324 nm by monitoring the hydrolysis of 15.6 μM *N*-succinyl-Ala-Ala-Pro-Phe-benzyl thioester (Bachem Biosciences) in the presence of 250 μM 4,4′-dithiodipyridine (Harper et al., 1981). Substrate and indicator were added as 100× concentrated stocks in *N*,*N*′-dimethylformamide to the preincubated enzyme−inhibitor complexes. The apparent inhibition constant (K_i*) was calculated as described previously for tight-binding ecotin−enzyme complexes [eq 1 in Seymour et al. (1994)]. All measurements were carried out in duplicate.

Structure Determination. Crystals of crab collagenase complexed to ecotin were grown as described (Tsu et al., 1994), with the further addition of 0.2 M sodium citrate to the precipitation reservoir prior to mixing with protein. The crystals were stabilized for data collection in an artificial mother liquor containing 40% PEG 4000, 0.2 M tricine (pH 8.5), and 0.2 M sodium citrate at a temperature of 17 °C. Crystals form in the trigonal space group $P3_221$ with cell dimensions $a = 89.11$ Å and $c = 291.55$ Å (Tsu et al., 1994). X-ray diffraction amplitudes were initially measured to 2.8 Å resolution on an R-AXIS IIC area detector mounted on a Rigaku RU-200 rotating anode generator. Data were obtained at ambient temperatures, at which the decay in intensity owing to radiation damage remained below 20% for a period of 16 h. Data from two crystals were integrated and scaled using the accompanying R-AXIS software package to yield an R_{merge} of 9.1% for 49 164 observations of 25 892 independent reflections with an $I/\sigma(I)$ of >1.0, representing 80% of the unique data to 2.8 Å resolution.

Table 1: Crystallographic Data for the Collagenase−Ecotin Complex

space group	$P3_221$
cell dimensions	$a = 89.11$ Å; $c = 291.55$ Å
tetramers/asymmetric unit	1
V_m (Å^3/Da)	4.2
solvent content	72%
highest resolution	2.3 Å
total observed	109 981
unique observed	41 721
% complete (2.3 Å/2.5 Å)	68%/78%
% complete (2.3−2.5 Å shell)	33%
% complete (2.5−2.6 Å shell)	51%
R_{merge}[a]	0.095
R_{cryst}[b]	0.189 (6.0−2.3 Å)
R_{free}	0.238 (6.0−2.3 Å)
rms bonds	0.011 Å
rms angles	2.8°
no. of waters	140

[a] $R_{merge} = (\sum_h\sum_i|\langle F_h\rangle - F_{hi}|)/(\sum_h F_h)$, where $\langle F_h\rangle$ is the mean structure factor magnitude of i observations of symmetry-related reflections with Bragg index h. [b] $R_{cryst} = (\sum_h\sum_i||F_{obs}| - |F_{calc}||)/(\sum|F_{obs}|)$ where F_{obs} and F_{calc} are the observed and calculated structure factor magnitudes, respectively.

The atomic coordinates of the tetrameric trypsin−ecotin complex (McGrath et al., 1994) were used as a search probe in the molecular replacement routines as implemented in XPLOR (Brunger, 1990). All atoms in the surface loops of trypsin, all side chain atoms beyond C^β in the trypsin core secondary structural elements, all ecotin side chain atoms beyond C^β in the regions of the primary and secondary sites, and all waters were deleted from the model. Intensity data in the range of 20.0−5.5 Å yielded a solution to the rotation function calculations. The strongest peak at 3.9σ above the mean was refined by Patterson correlation methods; this solution also persisted in calculations carried out using data in different resolution ranges. Translation function calculations were carried out in each of the space groups $P321$, $P3_121$, and $P3_221$; a solution at 11.6σ above the mean was obtained in space group $P3_221$ and was refined by rigid-body methods to a conventional crystallographic *R*-factor of 47.7% for data in the range of 6.0−2.9 Å. Further analysis of both self-rotation and cross-rotation functions gave no evidence for a second independent solution; with one tetramer per asymmetric unit, the Matthews coefficient is 4.2 Å^3/Da, indicating a very high solvent content of 72% (Table 1).

Crystallographic refinement in XPLOR (Brunger et al., 1987), including one round of molecular dynamics simulated annealing, was iterated with model building using FRODO (Jones, 1978) and CHAIN (Sack, 1988). Application of the noncrystallographic symmetry operator which relates the two halves of the collagenase−ecotin tetramer permitted rebuilding of only a single enzyme and inhibitor in the earlier stages. Refinement proceeded iteratively with and without the application of weighted NCS symmetry restraints as an additional empirical energy term. No real space density averaging was applied. Five rounds of positional and individual *B*-factor refinement iterated with rebuilding resulted in a nearly complete model with a crystallographic *R*-factor of 18.7% (6.0−2.8 Å) and an rms deviation in bond lengths of 0.011 Å.

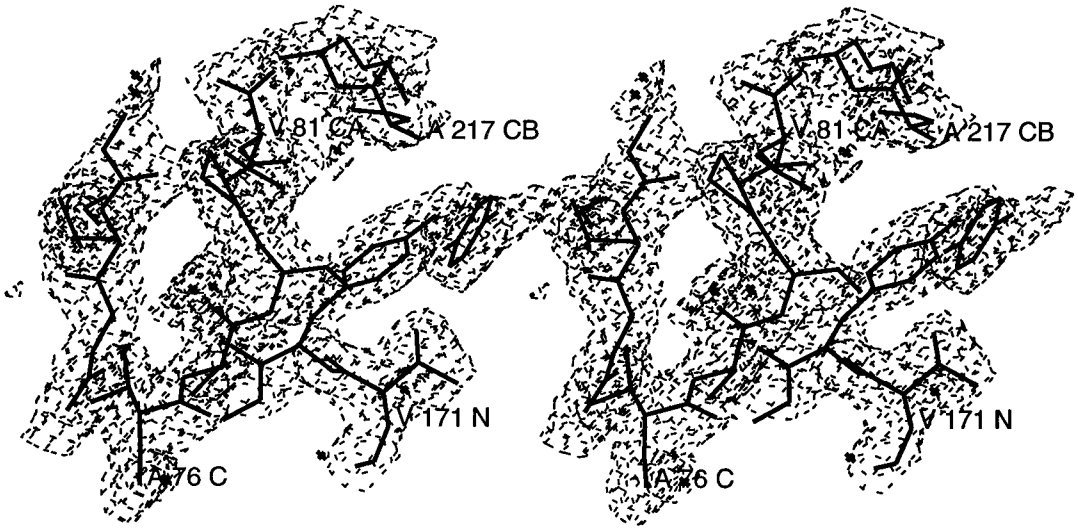


FIGURE 1: Stereoview of a simulated annealing OMIT map (Hodel et al., 1992) of the refined collagenase−ecotin structure in the region of the extended peptide binding site. The map is computed with coefficients $2F_o - F_c$ over the resolution range of 2.3−20 Å and is contoured at 1.0σ. A 10.0 Å sphere centered at ecotin Ser79 (the P5 inhibitor residue) was defined as the omitted region, and atoms in a 3 Å shell about this sphere were harmonically restrained to prevent artificial movements. Very tight harmonic restraints were applied to water molecules. Simulated annealing was carried out beginning at an initial temperature of 1000 K with cooling in increments of 25 K to a final temperature of 300 K. One hundred cycles of positional refinement were performed following the annealing procedures. The map was calculated using calculated amplitudes and phases derived from these coordinates, with all atoms inside the defined 10.0 Å sphere still omitted. The ecotin chain from residues Ala76 to Val81 (P8−P3) is shown from bottom to top, with residues in two collagenase segments comprising the extended peptide binding site at right. The electron density at left is from β-strand E of ecotin. This figure was produced using CHAIN (Sack, 1988).

Further data extending to 2.3 Å resolution, obtained on one additional crystal of a similar size, were collected on an RAXIS IIC area detector equipped with focusing mirrors at Molecular Structure Corp. (The Woodlands, TX). Data from this and several earlier crystals were reduced using the accompanying R-AXIS software to a final R_{merge} of 9.5% to 2.3 Å resolution [$I/\sigma(I) > 0.9$; Table 1]. Since data in the resolution shell extending between 2.3 and 2.5 Å resolution are only 33% complete [$I/\sigma(I) > 0.9$], we expect that the quality of the structure is better described as determined to 2.5 Å. Final rounds of refinement, rebuilding, and addition of 140 water molecules resulted in a crystallographic *R*-factor of 18.9% (6.0−2.3 Å; Table 1). The free *R*-factor (Brunger, 1992) following simulated annealing refinement of the final model is 23.8%, using a test set in which 10% of the data was removed. Computation of simulated annealing omit maps (Hodel et al., 1992) confirms a correct interpretation of several key portions of the model (Figure 1). The complete chain is traced for all four molecules in the asymmetric unit with the exception of the N-terminal four or five amino acids of each ecotin molecule, respectively. All side chains were modeled for each collagenase, while a small number in each ecotin monomer were disordered and included as alanines in the final model. The structure was analyzed using InsightII (Dayringer et al., 1986) running on a Silicon Graphics workstation. Superpositions were also carried out with the program OVRLAP (Rossmann & Argos, 1976). Atomic coordinates have been submitted to the Protein Data Bank at Brookhaven National Laboratory.

Model Building. The model of triple-helical collagen complexed to collagenase was constructed using the InsightII and Discover molecular modeling and energy minimization programs (Dayringer et al., 1986). Electrostatic calculations were performed using the program DELPHI (Honig et al., 1993) as implemented in the InsightII software.

RESULTS AND DISCUSSION

Overall Structure of the Tetrameric Complex. The tertiary structure of crab collagenase is, as expected, globally similar to those of the other chymotrypsin-like proteases. One hundred eighty-six of 226 α-carbons of collagenase superimpose upon their structural equivalents in rat trypsin with an rms deviation in position of only 0.87 Å. This degree of similarity is obtained among all pairs of these homologous enzymes (Greer, 1990; Perona & Craik, 1995). In this protein inhibitor complex, the side chains of the catalytic residues Ser195, His57, and Asp102 adopt positions identical to those observed in analogous serine protease structures, i.e., the trypsin−BPTI complex (Huber et al., 1974). The nucleophilic O^{γ} of Ser195 is in sub-van der Waals contact with the carbonyl carbon atom of the P1−P1′ Met84−Met85 bond of ecotin. At these positions, the conformation of the ecotin chain is also identical to that observed in other serine protease−protein inhibitor structures, including the ecotin−trypsin complex (McGrath et al., 1994), as is required for inhibition rather than cleavage to occur (Marquart et al., 1983).

The tetrameric collagenase−ecotin complex is composed of one inhibitor dimer and two enzyme molecules and, as observed in the trypsin-ecotin complex, is organized such that each protease contacts both ecotin monomers (Figure 2). This is accomplished by way of the secondary binding site, which comprises two structural elements of each protein. In each of the two secondary sites, one surface loop together with the C-terminal α-helix of the enzyme interact with two distal loops of the inhibitor. The primary and secondary binding sites of each ecotin monomer bind different collagenase molecules. The total surface area buried in these interactions is similar to that of the trypsin−ecotin complex (McGrath et al., 1994; Table 2). The modified jellyroll fold of the ecotin monomer is preserved in this complex, as is

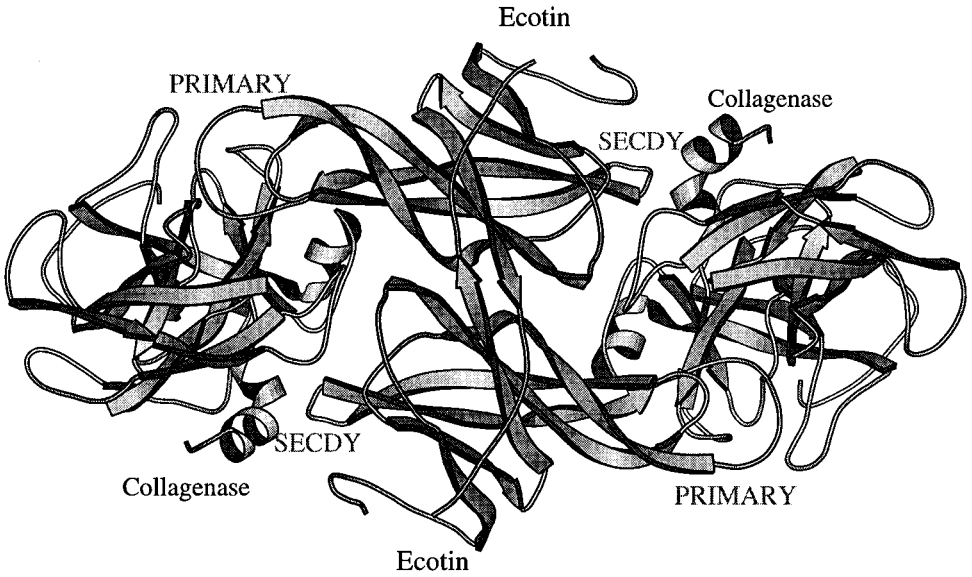


FIGURE 2: Schematic of the tetrameric collagenase−ecotin complex. The ecotin dimer (center) forms a network-linked tetramer with two collagenase molecules. The center two antiparallel β-strands comprise the majority of the ecotin dimer interface. The primary and secondary sites are indicated. This figure was produced with MOLSCRIPT (Kraulis, 1991).

Table 2: Comparison of Tetrameric Ecotin Complexes[a]

complex	number of contact residues		number of direct intermolecular hydrogen bonds	buried surface area (Å^2)
	Ecotin	Trypsin		
primary site	14	20	9	1900
secondary site	8	10	8	950
	Ecotin	Collagenase		
primary site	15	24	12	1950
secondary site	11	11	10	1060

[a] The criteria for contact residues are those making polar interactions (ionic and/or hydrogen bonding) or van der Waals interactions (3.8 Å or less). Intermolecular hydrogen bonds meet criteria of 3.4 Å or less between the two electronegative atoms.

the dimer interface comprising mainly the two long C-terminal β-strands which interact to form a two-stranded antiparallel β-ribbon.

Primary Binding Site. The primary binding loop of ecotin, spanning amino acids 78−88, makes interactions which extend over 11 enzyme subsites from S7 to S4′ (Figure 3). This interface between ecotin and collagenase is substantially larger than that formed in the ecotin−trypsin complex (McGrath et al., 1994), in which only the S3−S2′ sites were bound. Nine main chain hydrogen bonds are observed in the interface. Five of these occur in the S1 and S3 sites and are common in all known structures of chymotrypsin-like serine proteases complexed with peptidyl or protein inhibitors [see Tsu et al. (1997)]. Two more hydrogen bonds occur at the P5 and P6 positions. The P5-Pro80 carbonyl oxygen accepts a hydrogen bond from Ala217(A), and the P6-Ser79 amide acts as a donor in an interaction with Val171 (Figure 3). These hydrogen bonds form part of the extended interface which also includes a binding pocket for the P4-Val81 residue and an additional interaction between the side chain of P7-Ser78 and Tyr172 of the enzyme (Tsu et al., 1997).

On the leaving group side of the scissile bond, two additional main chain hydrogen bonds are formed by the P2′ residue, in an antiparallel β-sheet structure with Phe41 of the enzyme (Figure 3). Further, the main chain nitrogen atom of P4′-Pro88 of ecotin lies just 3.5 Å from the backbone amide oxygen of Met39, in a favorable hydrogen bonding orientation. It seems likely that the presence of an amide linkage at this position would lead to the formation of yet another main chain hydrogen bond, bringing the total number to 10. Thus, collagenase repeats a pattern on the leaving group side of the scissile bond of antiparallel β-sheet main chain hydrogen bonding interactions similar to that found on the amino-terminal side. In the ecotin−trypsin complex, only the hydrogen bond proximal to the scissile bond on the leaving group side is observed (McGrath et al., 1994). The absence of the other interactions arises as a consequence of a different conformation of the surface loop at residues 34−42 in that enzyme.

Secondary Binding Site. The secondary binding site is located some 20 Å from the primary area of interaction and is formed from the juxtaposition of two polypeptide segments of collagenase with two ecotin surface loops from the paired monomer. One segment of the protease chain bridges the two ecotin contact regions. Residues His91-Glu92-Asn93-Tyr94 forming an edge of a β-sheet in the secondary site are immediately N-terminal to the loop containing Phe97-Val98-Ile99 at one edge of the primary site. The other collagenase peptide segment is located close to the C terminus of the enzyme and forms an α-helix.

The secondary site features 10 direct intermolecular hydrogen bonds engaging both main chain and side chain groups, together with a large number of van der Waals interactions. Together, collagenase side chains His91, Glu92, and Asn93 make seven of these interactions with atoms of both ecotin loops at residues 67−70 and 108−113. The C-terminal helical segment of collagenase makes exclusively main chain hydrogen bonds at residues Tyr233, Asp236, and Phe237, each with Gly66 of ecotin. There are no buried water molecules observed at this interface.

It is anticipated that interactions in this secondary site will modulate and broaden the inhibitory profile of ecotin by providing an additional contact area not influenced by the primary target specificity. However, comparison of the sequences of trypsin and collagenase in this region shows

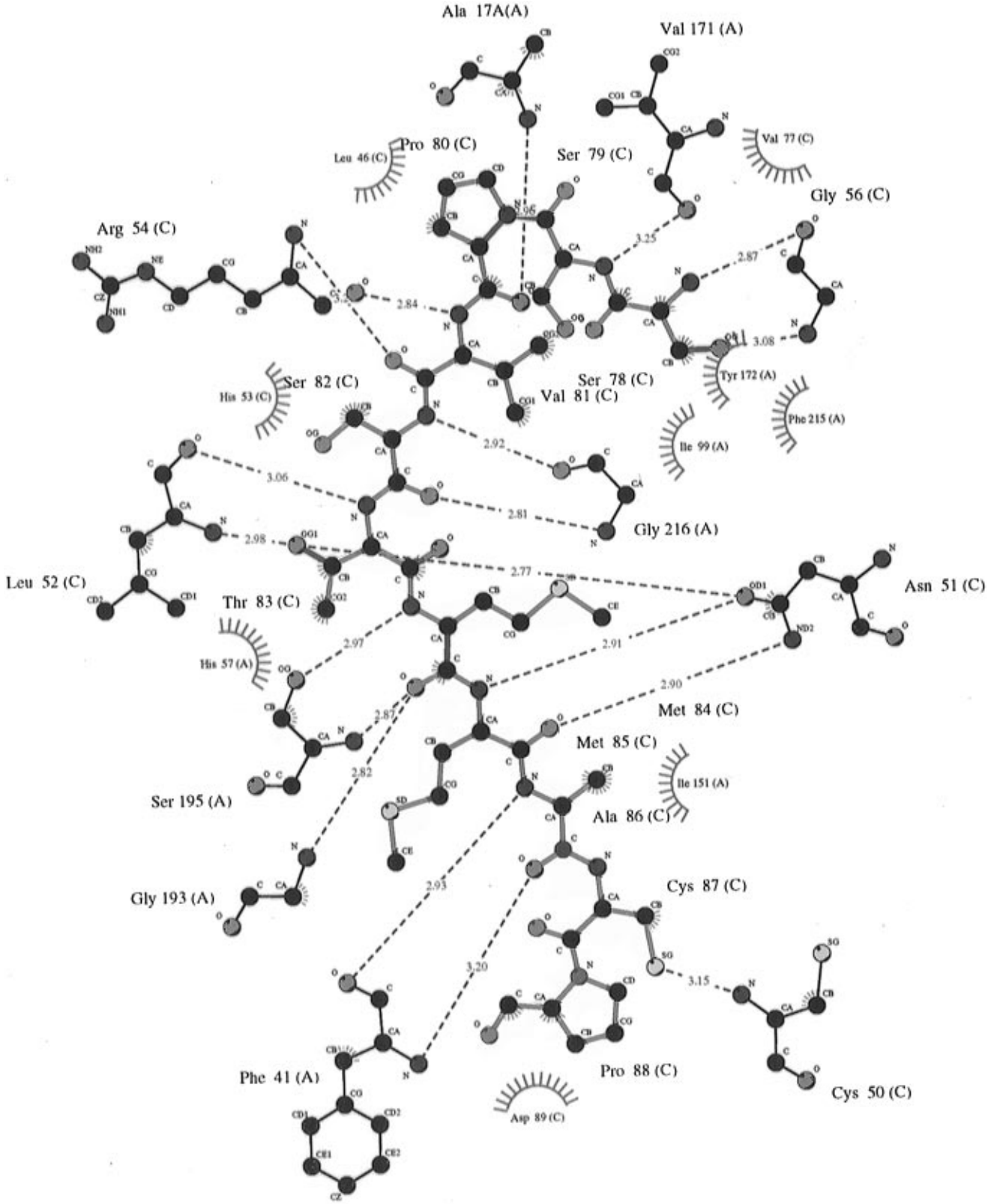


FIGURE 3: Interactions of the primary binding loop of ecotin from the P7−P4′ sites [residues Ser78(C)−Pro88(C)]. Collagenase residues are designated with an A, and ecotin residues are designated with a C. The nine intermolecular main chain hydrogen bonds described in the text are made by enzyme residues Val171(A), Ala217(A), Gly216(A), Ser195(A), Gly193(A), and Phe41(A). Intramolecular interactions made with the secondary binding site loop of ecotin [residues Cys50(C)−Gly56(C)] are also shown. Hydrogen bonds are indicated by dashed blue lines with the distances between the two electronegative atoms shown. van der Waals interactions are shown as green curved surfaces with the enzyme and inhibitor amino acids noted. The atoms are color-coded: nitrogen, blue; carbon, black; oxygen, red; and sulfur, yellow.

that, of the 12 amino acids present in the two enzyme segments 91−94 and 233−240, six are identical and two others are highly conservative substitutions (Grant et al., 1980). As might be expected, the secondary site interactions of the two complexes are consequently quite similar (Figure 4). Two additional hydrogen bonds are found in the collagenase complex relative to that of trypsin. In collagenase, Glu92 is substituted for Pro in trypsin, resulting in one

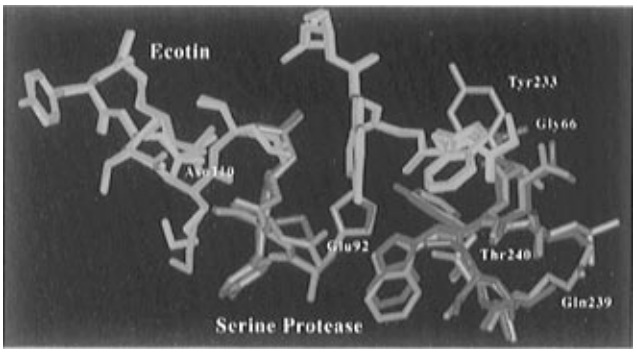


FIGURE 4: Superposition of the secondary binding sites in the ecotin−collagenase and ecotin−trypsin complexes. Enzyme residues in equivalent positions which differ significantly in chemical functionality are labeled (white-lettered label designations correspond to amino acids in collagenase). Yellow labels represent amino acids in ecotin. Collagenase amino acids are shown in orange and trypsin in red. The ecotin secondary site loops bound to collagenase and trypsin are shown in green and light blue, respectively.

additional contact. A second hydrogen bond is missing in the trypsin complex between ecotin Gly66 and the backbone carbonyl of enzyme residue Tyr233, owing to a slight bending away of the ecotin backbone at this position. The Tyr233 side chain may contribute indirectly to this difference as it provides van der Waals interactions with the aliphatic portion of the adjacent ecotin Glu65 side chain; in the trypsin complex, the equivalent enzyme amino acid is Asn233, and Glu65 is disordered beyond C^{β}. The remaining eight hydrogen bonds are identical in the two complexes. One additional difference relates to the solvent structure in the vicinity; in the trypsin complex, there are five water-mediated hydrogen bonds as compared to only two in the interface with collagenase (McGrath et al., 1994).

Ecotin Structural Flexibility. The structures of both the ecotin monomer and dimer differ markedly in this complex relative to their conformations bound to trypsin. There are large deviations in some parts of the structure, and identical conformations in other regions. The ecotin structural alterations apparently arise both from differences in the detailed structures of each protease target site and from a slightly altered spatial orientation of the two sites with respect to each other. It appears that many of the rearrangements occur as a consequence of the new P4−P7 contacts in the collagenase complex, since a portion of one ecotin β-strand (strand F; Figure 5a) directly molds itself to this more extended enzyme surface. Remarkably, however, the structural adaptation of the remainder of the molecule is not localized but instead encompasses substantial movements not only of the adjacent strand E but also of strands D and G of the other β-sheet of the modified jellyroll fold (McGrath et al., 1994, 1995). In the collagenase tetramer, the ecotin main chain in a surface loop (residues 90 and 91) on one edge of the primary site, which was disordered in the trypsin complex, is ordered in this complex and has been built into the final model.

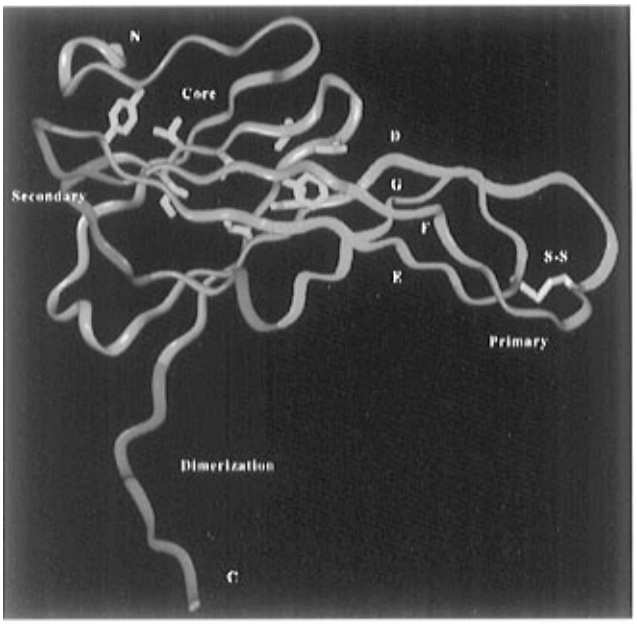
Two separate portions of the ecotin monomer maintain identical conformations in the two complexes. First, the structure of small segments of the two loops which make up the ecotin primary site is preserved (Figure 5a). These segments comprise the backbone atoms and disulfide bond of ecotin residues Ser82-Thr83-Met84-Met85-Ala86-Cys87 at the P3−P3′ sites and of residues Cys50-Asn51-Leu52 in the adjacent loop. The sole significant deviation in this region is at Ala86 in the P2′ site, where there is a divergence of 0.8 Å in the position of the carbonyl oxygen, possibly to facilitate formation of the hydrogen bond with Phe41 of collagenase (Figure 3). These nine amino acids define a maximal portion of the ecotin structure which must be maintained in a strict conformation, in order to preserve inhibitory function.

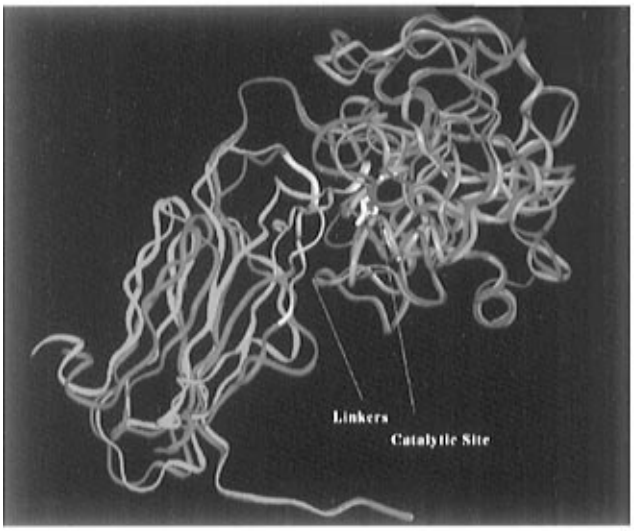
The second region of ecotin which is identical in structure in the two complexes comprises a large core domain consisting of 81 residues in three separate segments of chain: amino acids 7−43, 59−76, and 106−131 (Figure 5a). Equivalent backbone atoms of the 81 amino acids comprising these segments superimpose with an rms deviation in position of 0.31 Å. Ten core hydrophobic side chains in this part of the structure, each of which is fully buried in the interior, superimpose with a similar congruence of 0.25 Å. It is notable that the two loops which form the secondary binding site are part of the core domain and are not significantly different in structure in the two tetrameric complexes. Ecotin has a very large surface area and concomitantly small hydrophobic core for a protein of its molecular weight; only a single portion of one β-strand at amino acids 115−117, together with the 10 aforementioned side chains, is entirely solvent-inaccessible. No buried water molecules are found within the core in either monomer of the two complexes.

The relative orientations of the two fixed portions of the ecotin monomer differ substantially depending upon which enzyme is complexed. This difference arises from an exceptional plasticity of the connecting segments at amino acids 44−49, 53−58, 77−81, and 88−102 (Figure 5b). Large portions of each of the β-strands D−G (McGrath et al., 1994) of the ecotin-modified jellyroll fold are involved in the rearrangements. However, despite differences in the position of equivalent atoms ranging up to 6 Å, there is a remarkable conservation of the intramolecular antiparallel β-sheet hydrogen bonding arrangement. Twelve hydrogen bonds which bridge the pairs of interacting β-strands D and G and E and F (Figure 5a) are conserved in the two different ecotin conformers. This represents a large majority of the total number of these interactions; only three additional unique interstrand hydrogen bonds are found in the trypsin complex, and just a single additional unique interaction of this type exists when bound to collagenase.

Structural differences in the ecotin monomers are also found in a surface loop at residues 103−106, which forms a small additional dimer interface area when bound to collagenase. New van der Waals contacts between the Asp103 and Met106 side chains of each monomer are made, although the side chain of Asp103 beyond C^{β} is disordered as in the trypsin complex.

Differences in the conformations of the ecotin dimer in the two complexes are revealed by superimposing their structures based on the equivalences in position of all backbone atoms in the 81-amino acid core domain of one of the monomers (Figure 5c). This shows an angular divergence of several degrees propagating to positional differences of 5 Å at the secondary site of the opposing monomer. Clearly, not only the internal structure of the ecotin monomer but also the dimerization interface has substantial inherent capacity for structural deformation in response to protease binding. Two aspects of the dimer





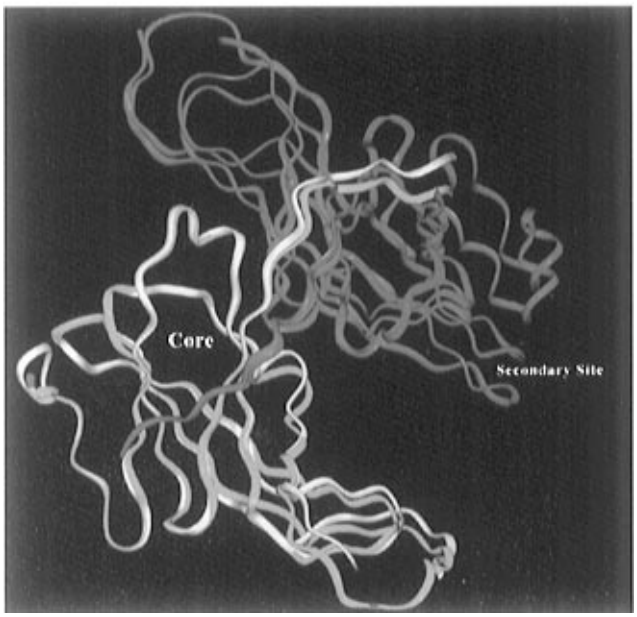


FIGURE 5: (a, top) Modular structure of ecotin derived from comparison of its conformation bound to collagenase and to trypsin. The flexible linkers corresponding to β-strands D−G (McGrath et al., 1995) are shown in green, and the parts of the structure which do not vary when bound to trypsin and collagenase are in orange. The primary site disulfide bond is in yellow and the area involved in dimerization in purple. Buried hydrophobic side chains in the core are drawn in blue. The secondary binding site of ecotin is shown in red. (b, middle) Superposition of trypsin (red) and collagenase (orange) showing the different structures of the β-strands (linkers) joining the primary binding site and the core region of ecotin. The linkers (β-strands E and F of ecotin) are shown in blue bound to collagenase and in yellow bound to trypsin. The catalytic residues of collagenase are shown in white. (c, bottom) Superposition of the structure of the ecotin dimers when bound to collagenase and to trypsin. The superposition was done on all backbone atoms of the core region of one of the monomers (green = ecotin bound to collagenase; yellow = ecotin bound to trypsin). The divergence in the structure of the dimer interface β-ribbon and of the position of the secondary site of the opposing monomer (purple = ecotin bound to collagenase; red = ecotin bound to trypsin) is evident.

interface are changed: (i) the conformation of the loop at positions 103−106, described above, and (ii) the orientation of the C-terminal arms, which diverge slightly at their base to produce substantial long-range atomic positional differences (Figure 5c) while retaining very similar detailed conformations.

A major driving force for the structural changes in the ecotin dimer may be a small difference in the spatial separation of the primary and secondary binding sites of collagenase relative to trypsin. Superposition of enzyme primary site main chain atoms at the oxyanion hole residues 193−195 and on the opposite side of the S1 pocket at residues 214 and 215 (rms deviation = 0.2 Å) shows that the secondary site C-terminal helical segment of trypsin is displaced from that of collagenase by 1.0−1.5 Å outward along the helix axis. This changed relative orientation of the two enzyme sites must, in and of itself, require some conformational adaptation on the part of ecotin. Since the primary and secondary enzyme sites bind different ecotin monomers, it is reasonable to expect that this difference in positioning of the two enzyme sites would be reflected in structural adaptation of the ecotin dimer interface. Of course, such adaptation could in general also occur in response to differences in the molecular surface of the secondary site, even if its position relative to the primary site is unchanged. In this regard, it will be of interest to determine the structures of other serine protease−ecotin complexes in which there is much less similarity in the amino acids comprising the secondary site. Such complexes might well reveal additional inherent capacity for deformation of the ecotin dimer interface.

The conformational rearrangements in ecotin, and the differences in the interaction surfaces in the trypsin and collagenase tetramers, prompted us to determine the inhibition constant for this complex. By monitoring the cleavage of a tetrapeptide thioester substrate, we find an apparent K_i of 0.5 nM (Figure 6), a value identical to the K_i of 0.5 nM measured against rat trypsin (Seymour et al., 1994). It is possible to speculate as to the structural origins of this identity; while an increased number of *intermolecular* hydrogen bonds and a greater buried surface area characterize

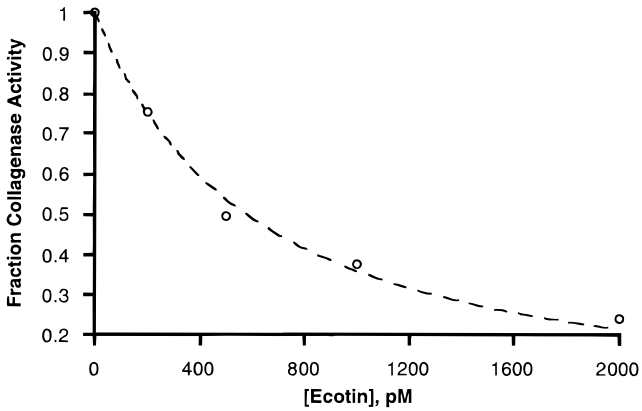


FIGURE 6: Inhibition of collagenase by ecotin. Reaction conditions were 50 mM Tris (pH 8.0), 100 mM NaCl, and 20 mM $CaCl_2$ at 25 °C (see Experimental Procedures). An apparent ecotin K_i of 510 ± 60 pM for collagenase was determined by two independent trials. The average value of the fractional enzyme activity remaining is shown for each of the experimental data points.

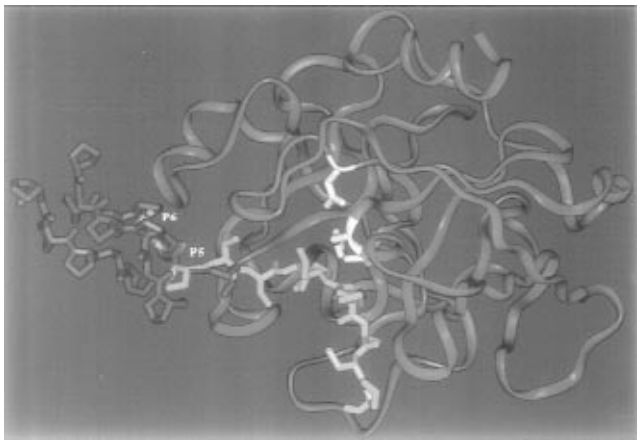


FIGURE 7: Superposition of a fragment of triple-helical collagen (red; PDB 1CLG) onto ecotin residues occupying the P5 and P6 subsites. The entire primary binding loop of ecotin from P7 to P4′ is shown in green; the collagenase catalytic residues (Ser195, His57, and Asp102) are in white.

the ecotin−collagenase complex, there are fewer *intramolecular* hydrogen bonds stabilizing the collagenase-bound conformer of ecotin. Therefore, energetic costs associated with maintaining this conformer could cancel the effects of additional favorable enthalpic interactions. The forthcoming crystal structure of uncomplexed ecotin should be informative with respect to this issue (Shin et al., 1993).

Structural features of ecotin likely to be common to the complexes formed with differing target proteases are the conformations of the two primary site loops, the conformation of the 81-amino acid core region (with the possible exception of the secondary site binding loops), and the presence of the antiparallel β-stranded dimer interface. In the remainder of the molecule, a range of β-sheet conformations, selected according to the requirements for inhibition of each particular serine protease, are likely to be sampled. Further, the N-terminal four or five residues of ecotin, which are disordered in both the trypsin and collagenase complexes, are recruited in the complex formed with chymotrypsin to provide additional secondary site interactions (C. Cambillau, personal communication). Given the known broad inhibition spectrum of ecotin (McGrath et al., 1995), it is unlikely that the full range of conformations accessible to this particular structural fold has as yet been elucidated. Continued crystallographic studies are needed to describe these conformations and to provide a foundation for understanding the underlying structural basis of the β-sheet plasticity.

Structural Determinants of Collagenolytic Specificity in Fiddler Crab Collagenase. The unique feature of crab collagenase relative to other well-characterized serine proteases is its ability to cleave a triple-helical collagen substrate. Clues to the structural basis of this specificity are revealed by the extended nature of the ecotin primary binding site interactions (Figure 3). No other chymotrypsin-like serine protease is known to possess so large a binding interface. The nature of the intermolecular interactions made at these subsites is also revealing. A total of nine main chain−main chain hydrogen bonds are made across the ten subsites from P6 to P2′. Further, the presence of a P4′-Pro in ecotin obscures visualization of a probable tenth such interaction with Met39 of collagenase (see below). Given the abundance of Gly and Pro residues in collagen, and the unimportance of the exposed side chains of the intervening amino acids for maintaining triple-helical structure, a preponderance of main chain hydrogen bonding contacts conforms to expectations for a protease−collagen complex. This observation provided an initial suggestion that the conformation of the primary binding loop of ecotin might mimic that adopted by a single polypeptide strand of collagen.

The polyproline (II) conformation of collagen chains possesses similar backbone $\phi-\psi$ angles to antiparallel and parallel β-sheets (Fraser et al., 1979). Thus, the antiparallel β-sheet hydrogen bonding between ecotin and collagenase, on both the N-terminal and leaving group sides of the scissile bond, indicates that the enzyme binding cleft is also well-suited to collagen binding. In particular, the conformation adopted by ecotin at the unique extended P5 and P6 sites can be exactly superimposed on that of a model of collagen derived from fiber diffraction data (Figure 7; PDB accession code 1CLG; rms deviation for the eight backbone atoms of these two amino acids = 0.16 Å). Apparently, the P5−P7 enzyme binding sites, including the main chain interactions at positions P5 and P6, form a surface which is highly complementary to the particular conformation of the polypeptide chains in collagen. The implication of this observation is that no melting or other conformational change of the triple-helix N-terminal to the P4 position may be required for collagen cleavage by fiddler crab collagenase.

The cleavage sites of crab collagenase within the α1- and α2-chains of type I collagen have been characterized in detail (Figure 8). Several primary cleavages occur between amino acids 775 and 795 at the expected three-fourths distances along each of the chains, although cleavage of the Leu587−Thr588 bond in α1 has also been detected. The local pattern of cleavage sites suggests that the observed or postulated main chain ecotin−collagenase hydrogen bonds at positions P4′, P2′, P1, P3, and P6 may play important roles in stabilizing a complex with collagen. Prolines at these positions are not found at any of the major sites of cleavage, even when the P1 and P1′ residues match strong enzyme preferences determined from cleavage of peptide substrates (Tsu et al., 1994; Pro is also disallowed at P1′ for all serine proteases). We surmise that some or all of these hydrogen bond donor functions, absent for Pro, may be important in stabilizing a collagen complex with the enzyme (the main chain interaction at P5 uses the substrate solely as acceptor

FIGURE 8: Cleavage sites of crab collagenase within the α1- and α2-chains of type I calf skin collagen [from Tsu et al. (1994)]. Sites were determined by Edman degradation analysis of one-fourth-length collagen fragments after cleavage reactions. Sequences of collagen are taken from Bornstein and Traub (1979). IC indicates the cleavage sites of vertebrate interstitial collagenase. Major sites of cleavage by crab collagenase are indicated with a 1 [45−60% of gel band as analyzed by Edman degradation; see Tsu et al. (1994)]; intermediate sites (15−40%) are designated with a 2, and minor sites (about 10%) are designated with a 3.

so that Pro is not excluded at this position). For example, the Phe−Leu bond from position 782 to 783 in the α2-chain is not cleaved despite strong preference for these amino acids in peptides, perhaps owing to disruptive Pro residues at P3 and P4′. A similar rationale appears to explain the poor cleavage of the Gln−Arg bond from position 788 to 789 in the α1-chain (another Gln−Arg bond nine amino acids upstream is cleaved very efficiently). Among sites of cleavage within this local region not disfavored by the presence of Pro at key positions, cleavage site selection parallels that observed toward short peptides, as described (Tsu et al., 1994). It is also of interest that the cleavage sites of vertebrate interstitial collagenase in both α1- and α2-chains possess Pro at the P3 position. This suggests that type I collagen−collagenase interactions in vertebrates may rely on a different set of interactions in the vicinity of the scissile bond. Together, these observations suggest that the conformation of the ecotin primary binding loop mimics that of the *cleaved* strand of triple-helical collagen.

While the main chain intermolecular hydrogen bonds observed in this ecotin−collagenase complex correlate with observed local collagen cleavage site preferences, it is not immediately evident why the triple helix at positions 587 and 775−790 is more collagenase-sensitive than are other domains of the molecule. The percentages of Pro and Hyp residues in these regions are not significantly lower than at other locations along the chains, as might be predicted for a more "relaxed" protease-sensitive domain. However, an interesting possibility is suggested from the high-resolution crystal structure of a collagen-like peptide (Bella et al., 1994), which showed that the triple helix is surrounded by a cylinder of hydration with extensive water−peptide hydrogen bonding interactions. Several closely spaced Gly-X-Y triplets in which X and Y are hydrophobic *amino* acids occur at the three-fourths cleavage site in both the α1- and α2-chains; such a pattern of hydrophobicity is unique to this region of the molecule. Hydrophilic *non-imino* acids have been suggested to participate in interchain water-mediated hydrogen bonding bridges on the basis of analysis of the collagen crystal structure, and it is possible that these interactions are important in the maintenance of triple-helical rigidity. Their relative absence at the three-fourths position may thus lead to a weakening of the structure and a thereby reduced free energy cost for the further deformation and melting likely to be required for cleavage. Such a pattern of hydrophobicity is not, however, observed in the vicinity of the cleaved Leu587−Thr588 bond in the α1-chain.

The structure of collagenase also reveals the presence of a large groove prominent on the leaving group side of the scissile bond, and extending across the active site to include the P3−P7 site interactions described above (Figure 9a,b). On the leaving group side, the side walls of this groove are formed from two surface loops spanning amino acids 34−42 and 74−80. These walls contain an abundance of charged amino acids. The floor of the groove in this region is formed from two segments of β-chain at amino acids 64−68 and 81−84 together with the Phe34 and Phe41 side chains at the edges of one surface loop. The pronounced groove is negatively charged as shown in Figure 9a. The negative electrostatic potential may interact with the overall +6 charge of the collagen cleavage site (Figure 8). Thus, a possible rationale for the highly acidic nature of the substrate-binding face of the enzyme is providing charge compensation for the basic residues present in both the α1- and α2-chains of collagen. While the β-strands and surface loops forming the proposed collagen binding site are present in analogous positions in all chymotrypsin-like serine proteases (Perona & Craik, 1995), their detailed conformations differ, and no other enzyme of known structure possesses a well-defined groove in this region. Strikingly, the width of the groove on the leaving group side of the scissile bond exactly accommodates the 14 Å diameter collagen triple helix (Figure 9b). This strongly suggests a function in substrate binding and in appropriately orienting the scissile strand into the catalytic site. The highly acidic overall p*I* of the enzyme, calculated to be 3.9 from the amino acid sequence, further suggests the possibility that electrostatic guidance may play a role in facilitating docking of the two proteins.

We have used the observed binding of the ecotin primary loop over the P7−P4′ sites, and the presence of the large groove, as the basis for construction of a model for the collagenase−collagen complex (Figure 10). This model is specific for scission of the Gln779−Arg780 bond of the α1-chain as the first cleavage event, after which it is expected that cleavages of the other two strands can be readily accomplished without bringing the full power of the enzyme to bear in binding and deforming the triple helix. We used the model for the $(PPG)_{10}$ structure derived from fiber-diffraction data (PDB accession code 1CLG) as a starting point for this modeling work, with the initial assumption (justified above) that the observed conformation of ecotin residues 78−88 (P4′-P7 sites) is an accurate model for the cleaved strand of the triple helix. It is then immediately evident that steric constraints arising from surrounding regions of the protein require that this strand be melted out of the native triple helix to a substantial degree.

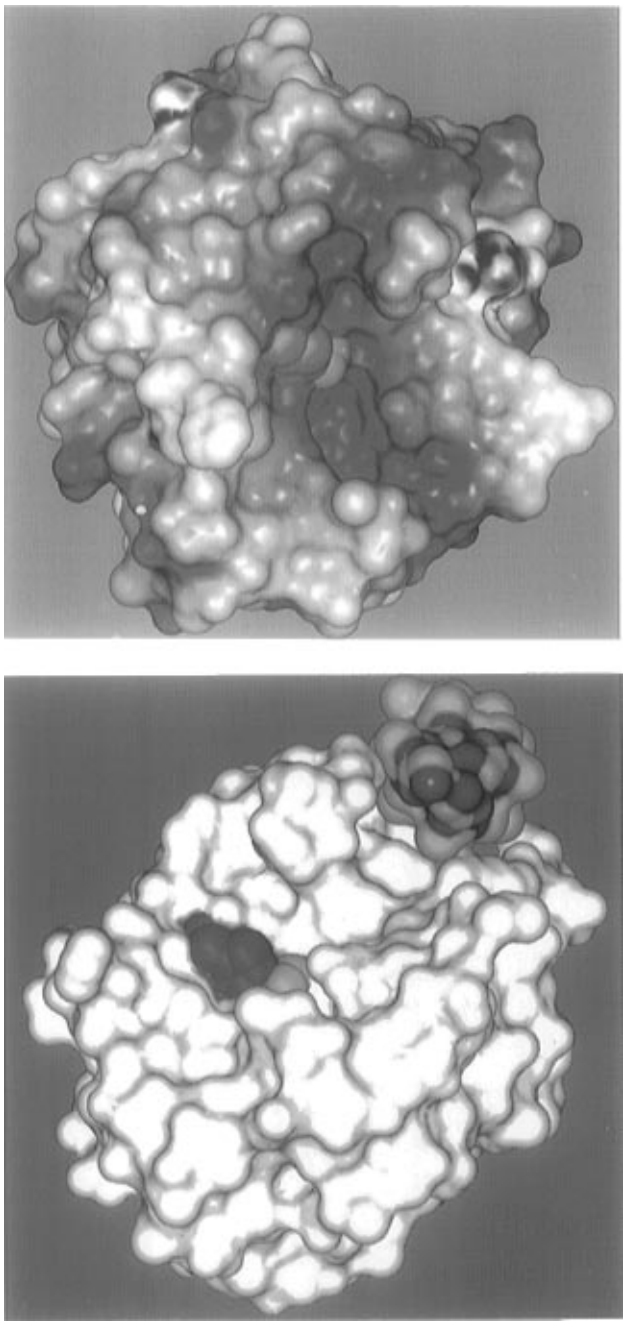


FIGURE 9: (a, top) Surface of crab collagenase with electrostatic potentials calculated using the program DELPHI. His57 is shown in blue, and Ser195 is in orange. Negative contours are in red and predominate owing to the highly acidic p*I*. In this view, the pronounced groove which binds collagen runs from the center-bottom of the structure to the center-right of the structure. Collagen residues N-terminal to the scissile bond bind at center-bottom, and the leaving group side of the substrate (C-terminal direction) is positioned at center-right. (b, bottom) Preliminary docking of triple-helical collagen in the collagenase groove. Shown is an idealized collagen helix with amino acids truncated to alanine, fit to the portion of the groove on the leaving group side of the scissile bond. Ser195 is in orange and His57 in purple at the center of the enzyme. This orientation is similar to that in panel a except that the enzyme is rotated roughly 45° about an axis parallel to the groove. In this view, the groove runs directly across the center of the molecule.

Substitution of the ecotin P4′−P7 amino acids with those surrounding the Gln779−Arg780 bond of the collagen α1-chain (Figure 8) shows the absence of any conflicting steric clashes with these amino acids. All of the collagenase binding sites readily accommodate the required collagen amino acids with little or no rearrangement of side chain groups and no required repositioning of main chain atoms. Inspection of the amino acids surrounding the other known cleavage sites (Figure 8) similarly shows an absence of steric hindrance and, in general, good complementarity with the corresponding enzyme subsites. Next, we note that extension of the cleaved collagen strand along the surface loop defined by enzyme residues 34−41 readily permits formation of several additional main chain−main chain hydrogen bonds C-terminal to the scissile bond. These interactions are not observed in the ecotin complex but are easily modeled without any rearrangement of the enzyme groups. The new interactions are at positions P4′ (the ecotin P4′-Pro backbone nitrogen is already observed to be adjacent to the carbonyl oxygen of Met39) and P6′. At P6′, the main chain amide of Asp38 of collagenase is well-positioned to accept a hydrogen bond from the collagen substrate. The cleaved strand of collagen may thus make as many as 11 exclusively main chain−main chain hydrogen bonds with the enzyme.

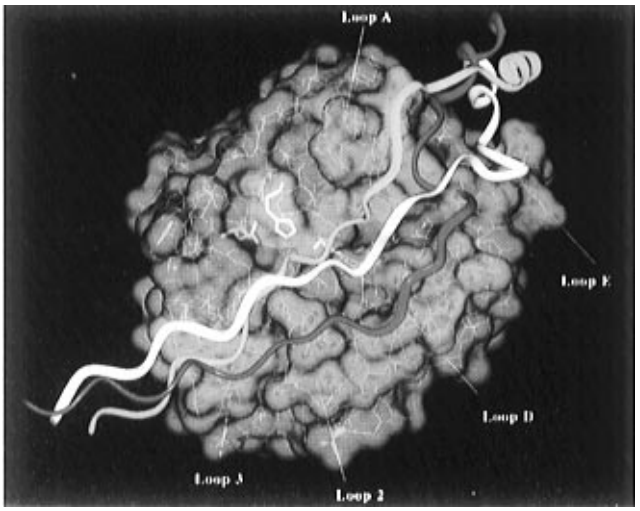


FIGURE 10: Connolly surface representation of crab collagenase with a modeled ribbon representation of the collagen substrate. The scissile strand is in green (matching coordinates of the primary binding loop of ecotin), and the enzyme catalytic residues are in white. The nomenclature for the surface loops follows Perona and Craik (1995). The loops help to determine subsite preferences among a number of enzymes in the chymotrypsin-like serine proteases and in collagenase adopt conformations which generate the collagen-binding groove. Loop A contains amino acids 34−41. Loop D contains amino acids 143−149. Loop E contains amino acids 74−80. Loop 2 contains amino acids 217−225. Loop 3 contains amino acids 169−174. The model has been energy minimized using the Discover force field.

Since the surface loop at residues 34−41 (loopA, Figure 10) forms one wall of the prominent groove on the leaving group side of the enzyme, making the proposed hydrogen bonds at P4′ and P6′ places the cleaved strand well above the floor of the groove. Modeling of the collagen triple helix into this groove shows that only two of the three strands are able to interact with the enzyme, as the third strand runs across the top of the other two and makes no direct contacts (Figure 10). It appears likely, then, that it is the strand of collagen which is cleaved which makes no enzyme contacts in this part of the groove. This topological feature of the complex constrains the rotational freedom of the collagen about its long axis; precise orientation of the two noncleaved strands with the floor of the groove is needed to position the P6′−P2′ residues of the cleaved strand for hydrogen bonding with the enzyme surface loop forming one wall.

Further, since type I collagen possesses two α1-chains and one α2-chain, an inherent asymmetry arises upon binding because of the asymmetry of the enzyme surface. The two noncleaved strands (either α1 and α2 as in the present model or two α1-chains in the event that α2 is cleaved first) will be placed in different environments. This provides the potential for additional constraints on the topology of productive complexes. However, prediction of the relative positions of the uncleaved α1- and α2-chains in this model, or indeed of whether α2 might be cleaved first, would require a precise definition of steric constraints on the sizes of amino acids occupying specific subsites in the *uncleaved* strands. There are insufficient constraints on the conformations of these strands to permit such predictions at this time.

The orientational freedom of collagen about its long axis is also constrained on the N-terminal side of the scissile bond by the observed main chain hydrogen bonds at the P5 and P6 positions. As noted above, the conformation of the ecotin backbone at these positions exactly matches that of the native collagen triple helix (Figure 7), and the model thus incorporates a regular triple-helical structure up to and including the P5 substrate residue. This includes the retention of an intramolecular hydrogen bond in which the P5 amide nitrogen is the donor group (the P5 amide oxygen accepts a hydrogen bond from the enzyme; Figure 3). Disruption of the intramolecular collagen hydrogen bonding then occurs beginning at the P4 position. On the C-terminal side, the model predicts that the triple helix is disrupted over approximately 13 amino acids following the P1 residue. This arises as a consequence of steric constraints from the groove in this region as well as from the extended binding site of the scissile strand to P6′. However, the disruption of the regular triple helix occurs two to three amino acids inside of the most distal predicted contact (residue P15′), a distance very similar to that observed on the N terminus. Because there is one intramolecular hydrogen bond per residue in collagen, this model predicts the disruption of some 19 or 20 of these stabilizing contacts in forming the enzyme complex. Much of this energetic cost is presumably compensated for by the interactions of the scissile and the two nonscissile strands with enzyme groups. Additionally, the two uncleaved strands appear to be able to maintain substantial intramolecular interactions (see below) which may include retention of some hydrogen bonds. The probable limits on the extent of melting define the axis of collagen as it enters and leaves the extended enzyme binding site and suggest that its path through space is bent by approximately 40° when bound to collagenase (Figure 10).

The collagenase structure also suggests a possible path for the two noncleaved strands across the enzyme surface. A surface loop at amino acids 143−149 (loop D; Figure 10) forms a shallow bowl bridging one edge of the extended peptide binding site (residues 216−219, loop 2; Figure 10) and the groove on the leaving group side. Beginning at the N-terminal edge, one collagen strand (Figure 10, red) can be readily modeled to extend across this surface and to emerge in the distal groove where it rejoins the cleaved strand to again form the triple-helical conformation. This strand maintains contact with the enzyme throughout the full length of 19 or 20 amino acids over which the helix is disrupted. Owing both to steric constraints and to the orientation of the third strand of the helix as it binds to the P5 site, it appears that this third chain of collagen (Figure 10, white) can make no direct interactions with the enzyme near the N-terminal side of the substrate. It has thus been modeled to maximize intramolecular interactions with the directly bound uncleaved strand (Figure 10, red). It seems necessary, however, that this polypeptide interact closely with the groove on the leaving group side of the enzyme, where it is the scissile strand which makes no direct contacts.

The model demonstrates the suitability of the large groove for binding collagen and describes likely constraints on the path of this substrate based on its known uncomplexed structure, on the experimental data correlating cleavage sites with the observed ecotin conformation, and on the presence of enzyme surface features. It may be easily tested by mutational analysis, as it should be possible to disrupt collagen binding in the groove while retaining high activity toward short peptide substrates. It also serves as a useful guide for predicting likely characteristics of the collagen binding sites of mammalian interstitial collagenases. These enzymes are composed of a thermolysin-like catalytic domain and a four-bladed propeller-like hemopexin domain, neither of which possesses surface features resembling the groove described here. Even the full-length porcine synovial collagenase (Li et al., 1995) apparently reveals no clues to the molecular basis for the collagenolytic specificity. In this structure, the two domains are connected by a proline-rich linker region of poorly defined structure. Thus, it may well be that a defined collagen binding groove is generated only by a substrate-induced conformational change which brings the two parts of the molecule together in a specific manner. Definition of this groove is key to structure-based enzyme inhibition strategies which might block tumor invasion, as targeting this particular region would allow for specificity against other enzymes possessing essential metalloprotease functions. Because the structure of collagen is conserved, however, it is conceivable that a targeted drug design strategy based on the shape of the crab collagenase binding groove could be implemented.

Versatility of Substrate Preferences in the Chymotrypsin-like Serine Proteases. The ability of invertebrate chymotrypsin-like serine proteases to cleave triple-helical collagen shows the remarkable versatility of the bilobal β-barrel structural scaffold. This scaffold and the associated catalytic side chains are very highly conserved in all the enzymes of this family. In spite of this, an inherent capacity for diversity in substrate specificity is nonetheless built in as allowed variability in the conformations of surface loops (Perona & Craik, 1995). Some of the enzymes in the family, including trypsin, chymotrypsin, and elastase, possess substantial specificity only at the primary substrate binding site (P1). Others such as enteropeptidase can possess substantial subsite preferences (Lavallie et al., 1993), and structure-based sequence alignments have implicated surface loops at positions 95−100 and 59−65 in providing key residues for recognition of an Asp_4-Lys substrate sequence. The structure of crab collagenase now allows the development of this theme in a unique manner: five different surface loops collaborate to generate a pronounced groove which appears to be eminently suitable for binding the triple-helical collagen substrate. Our model, generated on the basis of the conformation of the ecotin primary loop and the known native collagen structure, clearly shows the suitability of this groove in providing a large number of binding interactions to offset the needed melting of the helix.

While structure-based sequence alignments are a powerful tool for predicting function for newly isolated enzymes belonging to a particular family, use of this tool is hampered in this case by a lack of required functional data for already-known enzymes. Establishing definitively that a chymotrypsin-like serine protease cleaves type I collagen requires that the appearance of the signature one-fourth- and three-fourths-length fragments be monitored (Tsu et al., 1994; Tsu & Craik, 1996). However, these data are available only for crab collagenase and not for apparently closely related enzymes isolated from a range of other organisms, including humans. This prevents elucidation of conserved sequence differences which could allow prediction of the presence of similar grooves in other enzymes of the family. However, among enzymes of known structure which have been tested in this assay (Tsu et al., 1994), we note that only crab collagenase both cleaves type I collagen and possesses the marked groove described here. Human neutrophil elastase, which is known to cleave type III collagen, possesses a much less well-defined surface groove. Possibly, the type III collagen helix is deformed at a lower free energy cost than is type I and thus does not require interactions with an enzyme binding site that are as extensive. A marked surface groove may be unlikely for the numerous gelatinases as well, since the structure of gelatin (denatured collagen) is known to not be tightly wound. Definitive one-fourth and three-fourths fragment generation from type I collagen is needed to avoid misclassifying candidate proteases as true collagenases. Crystallographic studies on such newly identified enzymes might then give substantial insight into which particular aspects of the binding groove are most important in providing the collagenolytic specificity.

ACKNOWLEDGMENT

We are grateful to Dr. Joseph Ferrara of Molecular Structure Corp. for assistance in data collection and processing and to Jennifer Harris for helpful discussions and critical reading of the manuscript.

REFERENCES

Bella, J., Eaton, M., Brodsky, B., & Berman, H. M. (1994) *Science 266*, 75−81.
Bode, W., Reinemer, P., Huber, R., Kleine, T., Schnierer, S., & Tschesche, H. (1994) *EMBO J. 13*, 1263−1269.
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*, 106−110.
Bornstein, P., & Traub, W. (1979) in *The Proteins* (Neurath, H., & Hill, R. L., Eds.) Vol. IV, pp 412−632, Academic Press, New York.
Brunger, A. T. (1990) *Acta Crystallogr. A 46*, 46−57.
Brunger, A. T. (1992) *Nature 355*, 472−474.
Brunger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458−460.
Chung, C. H., Ives, H. E., Almeda, S., & Goldberg, A. L. (1983) *J. Biol. Chem. 258*, 11032−11038.
Dayringer, H., Tramontano, A., Sprang, S., & Fletterick, R. J. (1986) *J. Mol. Graphics 4*, 82−87.
Faber, H. R., Groom, C. R., Baker, H. M., Morgan, W. T., Smith, A., & Baker, E. N. (1995) *Structure 3*, 551−559.
Fraser, R. D. B., MacRae, T. P., & Suzuki, E. (1979) *J. Mol. Biol. 129*, 463.
Grant, G. A., Henderson, K. O., Eisen, A. Z., & Bradshaw, R. A. (1980) *Biochemistry 19*, 4653−4659.
Greer, J. (1990) *Proteins*: *Struct., Funct., Genet. 7*, 317−334.
Harper, J., Ramirez, G., & Powers, J. (1981) *Anal. Biochem. 118*, 382−387.
Hirose, T., Patterson, C., Pourmotabbed, T., Mainardi, C. L., & Hasty, K. A. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 2569−2573.
Hodel, A., Kim, S.-H., & Brunger, A. T. (1992) *Acta Crystallogr. A 48*, 851−858.
Honig, B., Sharp, K., & Yang, A.-S. (1993) *J. Phys. Chem. 97*, 1101−1107.
Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J., & Steigemann, W. (1974) *J. Mol. Biol. 89*, 73−101.
Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268−272.
Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946−950.
LaVallie, E. R., Rehemtulla, A., Racie, L. A., DiBlasio, E. A., Ferentz, C., Grant, K. L., Light, A., & McCoy, J. M. (1993) *J. Biol. Chem. 268*, 23311−23317.
Lecroisey, A., Gilles, A.-M., de Wolf, A., & Keil, B. (1987) *J. Biol. Chem. 262*, 7546−7551.
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. (1995) *Structure 3*, 541−549.
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*, 375−377.
Marquart, M., Walter, J., Deisenhofer, J., Bode, W., & Huber, R. (1983) *Acta Crystallogr. B 39*, 480−490.
Matthews, B. W., Jansonius, J. N., Colman, P. M., Schoenborn, B. P., & Dupourque, D. (1972) *Nature 238*, 37−41.
McGrath, M. E., Hines, W. M., Sakanari, J. A., Fletterick, R. J., & Craik, C. S. (1991) *J. Biol. Chem. 266*, 6620−6625.
McGrath, M. E., Erpel, T., Bystroff, C., & Fletterick, R. J. (1994) *EMBO J. 13*, 1502−1507.
McGrath, M. E., Gillmor, S. A., & Fletterick, R. J. (1995) *Protein Sci. 4*, 141−148.
Pal, G., Sprengel, G., Patthy, A., & Graf, L. (1994) *FEBS Lett. 342*, 57−60.
Perona, J. J., & Craik, C. S. (1995) *Protein Sci. 4*, 337−360.
Rich, A., & Crick, F. H. C. (1961) *J. Mol. Biol. 3*, 483.
Rossmann, M. G., & Argos, P. (1976) *J. Mol. Biol. 105*, 75−95.
Sack, J. S. (1988) *J. Mol. Graphics 6*, 224−225.
Sanchez-Lopez, R., Alexander, C. M., Behrendtsen, O., Breathcah, R., & Werb, Z. (1993) *J. Biol. Chem. 268*, 7238−7247.
Sellos, D., & Van Wormhoudt, A. (1992) *FEBS Lett. 309*, 219−224.
Seymour, J. L., Linquist, R. N., Dennis, M. S., Moffat, B., Yansura, D., Reilly, D., Wesswinger, M. E., & Lazarus, R. A. (1994) *Biochemistry 33*, 3949−3958.
Shin, D. H., Hwang, K. Y., Kim, K. K., Lee, H. R., Lee, C. S., Chung, C. H., & Suh, S. W. (1993) *J. Mol. Biol. 229*, 1157−1158.
Spurlino, J. C., Smallwood, A. M., Carlton, D. D., Banks, T. M., Vavra, K. J., Johnson, J. S., Cook, E. R., Falvo, J., Wahl, R. C., Pulvino, T. A., Weendoloski, J. J., & Smith, D. L. (1994) *Proteins 19*, 98−109.
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*, 119−123.
Tsu, C. A., & Craik, C. S. (1996) *J. Biol. Chem. 271*, 11563−11570.
Tsu, C. A., Perona, J. J., Schellenberger, V., Turck, C. W., & Craik, C. S. (1994) *J. Biol. Chem. 269*, 19565−19572.
Tsu, C. A., Perona, J. J., Fletterick, R. J., & Craik, C. S. (1997) *Biochemistry 36*, 5393−5401.
Woessner, J. F., Jr. (1991) *FASEB J. 5*, 2145−2154.

BI9617522