Structural Basis for the Broad Substrate Specificity of Fiddler Crab Collagenolytic Serine Protease 1[†]

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

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: Crab collagenolytic serine protease 1 efficiently cleaves peptide bonds directly C-terminal to basic, polar, and hydrophobic amino acids. The crystal structure of this enzyme complexed to the protein inhibitor ecotin at 2.5 Å resolution reveals a large primary binding pocket punctuated on one wall by the side chain of aspartate-226. Removal or relocation of this negatively charged group by site-directed mutagenesis generates variant enzymes which retain very high activities toward selected substrates. Full retention of activity toward hydrophobic substrates in collagenase D226G is accompanied by a 10−100-fold reduction in k_{cat}/K_m toward basic residues. In contrast, restoration of the negative charge in a trypsin-like position in collagenase D226G/G189D regenerates nearly full activity toward basic substrates while introducing a 5-fold decrease in k_{cat}/K_m toward hydrophobic amino acids. These results imply that the collagenase S1 pocket has multiple distinct binding sites for different amino acid side chains, a suggestion supported by molecular modeling studies based on the crystal structure. The ease of specificity modification in the primary binding site of this serine protease parallels similar observations with the bacterial enzymes α-lytic protease and subtilisin, and stands in sharp distinction to the extensive mutagenesis required to alter specificity in trypsin.

Collagenolytic serine protease 1 (EC 3.4.21.32) isolated from the hepatopancreas of the fiddler crab, *Uca pugilator*, was the first serine protease found to be capable of cleaving native type I triple-helical collagen (Eisen et al., 1973). Serine collagenases have since been isolated from a variety of organisms and are thought to be primarily involved in the digestion of foodstuffs (van Wormhoudt et al., 1992; Eisen et al., 1973). In addition to its eponymous activity, crab collagenase also possesses significant trypsin, chymotrypsin, and elastase-like substrate specificities and is the most efficient serine protease known in the hydrolysis of P1-Gln and P1-Leu amide substrates (Grant & Eisen, 1980; Welgus et al., 1981; Tsu et al., 1994). Preferences in the cleavage of peptide bonds within a defined domain of collagen mirror the enzyme's specificity toward small peptidyl substrates (Tsu et al., 1994). Kinetic analysis also demonstrates that a large fraction of the catalytic efficiency of crab collagenase is derived from binding of the extended substrate residues at positions P2−P4 (Tsu & Craik, 1996), a property which may reflect in part the collagenolytic activity of the enzyme. The ability of collagenase, but not trypsin, chymotrypsin, or other homologs, to cleave triple-helical collagen arises from unique extended substrate binding sites in the former enzyme (Perona et al., 1997).

The amino acid sequence of fiddler crab collagenase identifies it as a member of the chymotrypsin-like serine proteases, as it maintains approximately 35% sequence identity with the mammalian trypsin, chymotrypsin, and elastase enzymes (Grant et al., 1980; Tsu & Craik, 1996). In this protease family, the presence of Asp189, Gly216, Gly226, and Ser/Thr190 amino acids in the S1 site correlates with exclusive specificity toward P1-Lys/Arg substrates (Greer, 1990; Perona & Craik, 1995). Crab collagenase maintains the conservation of these residues with one interesting exception. The negative charge at the base of the pocket is repositioned so that amino acids Asp189 and Gly226 of trypsin are altered to Gly189 and Asp226 in collagenase (Grant et al., 1980). The reconfiguration of binding pocket geometry does not suggest a basis for broad specificity; modeling of the collagenase sequence on the trypsin scaffold predicts that hydrophobic P1 substrates should interact unfavorably with Asp226. Further, a trypsin variant in which Asp189 and Gly226 are interchanged (trypsin D189G/G226D) retains a substrate specificity profile exclusively toward P1-Lys and Arg amino acids and maintains accessibility of Asp226 to substrate (Perona et al., 1993). Clearly, the early model (Stroud, 1974), which proposed that specificity depends only on a small number of amino acids in the S1 pocket, cannot account for the large differences in efficiency toward hydrophobic substrates exhibited by collagenase and trypsin D189G/G226D. Substrate specificity in trypsin, collagenase, and possibly other members of the chymotrypsin-like serine protease family must arise instead from a more extensive set of determinants. Supporting this view is the observation that exchange of amino acids in three distal polypeptide segments is required

[†] 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 Biochemistry, University of Colorado, Boulder, CO 80309.

[||] Present address: Department of Chemistry and Interdepartmental Program in Biochemistry and Molecular Biology, University of California, Santa Barbara, CA 93106.

[⊥] Department of Biochemistry and Biophysics.

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

S0006-2960(96)01753-9 CCC: $14.00

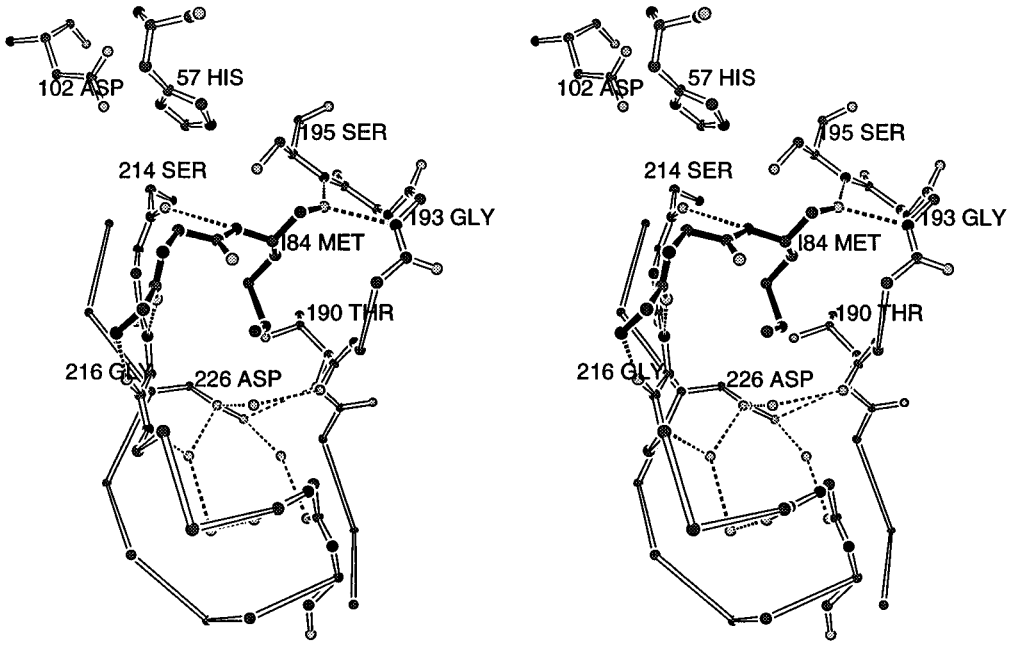

FIGURE 1: Stereoview of the S1 site of crab collagenase (open bonds), shown interacting with the P1−P3 residues of ecotin (filled bonds). Hydrogen bonding interactions are shown as dotted lines. The five main chain enzyme−inhibitor hydrogen bonds comprise the oxyanion hole interactions at upper right and the antiparallel β-sheet-like contacts at left. Nitrogen, carbon, and oxygen atoms are black, darkly shaded, and lightly shaded, respectively. Isolated spheres represent the positions of buried water molecules.

to alter the specificity of trypsin to that of chymotrypsin (Hedstrom et al., 1992, 1994). In trypsin and chymotrypsin, the distal segments function by determining the conformation of Gly216, which operates as a specificity determinant by accurately aligning substrates with respect to the catalytic residues (Perona et al., 1995).

We present here a detailed description of the S1 specificity pocket of crab collagenase interpreted on the basis of an electron density map of the collagenase−ecotin complex at 2.5 Å resolution (Perona et al., 1997). We find that the pocket is enlarged relative to that of trypsin, owing to the insertion of several amino acids at one edge. The hydrophobic P1-methionine of ecotin does not enter the pocket and remains 5 Å distant from the solvent-accessible Asp226 residue. The role of Asp226 as a primary substrate determinant was further explored via site-directed mutagenesis experiments. Kinetic analysis of mutant enzymes corroborates the results of structure-based modeling and suggests that the broad specificity of collagenase is derived from the ability of the S1 site to accommodate different amino acid side chains in distinct positions.

EXPERIMENTAL PROCEDURES

Production of Site-Directed Mutants of Collagenase. Site-directed mutagenesis of the collagenase gene was achieved in the PsFC (Tsu & Craik, 1996) vector, using the uracil-laden single-stranded DNA method (Kunkel, 1985). Recombinant collagenase was expressed in *Saccharomyces cerevisiae* using the PyFC shuttle vector and was purified as described (Tsu et al., 1994; Tsu & Craik, 1996). Enzyme purity was determined by Coomassie-stained SDS−PAGE[1] and was found to be ≥95% in all cases (Laemmli, 1970).

Enzyme Kinetics. Steady state kinetic assays were run in 50 mM Tris (pH 8.0), 100 mM NaCl, 20 mM $CaCl_2$, 0−2% DMF, and 0−9.8% Me_2SO at 25 °C. Benzyl thioester reactions included 250 μM dithiodipyridine (Harper et al., 1981). The concentration of active collagenase was determined by active site titration using the active site titrant MUTMAC (Sigma) (Jameson et al., 1973). Z-Y-Sbzl was purchased from Enzyme Systems Products. All other substrates were purchased from Bachem Biosciences. Spectroscopic measurements were carried out as described (Corey & Craik, 1992). Initial rates were fit directly to the Michaelis−Menten equation (Fersht, 1985).

Collagen Cleavage Assays. Calf skin collagen was from US Biochemical. Collagen cleavage assays were carried out as described (Tsu et al., 1994; Tsu & Craik, 1996) except that the microfiltration step was omitted.

X-ray Crystallography. Crystal structure determination of the collagenase−ecotin complex is described in Perona et al. (1997). Molecular modeling of alternative substrates was carried out using the program InsightII (Dayringer et al., 1986) running on a Silicon Graphics workstation.

RESULTS

Structure of the Primary S1 Binding Site of Fiddler Crab Collagenase. As in trypsin and chymotrypsin, the S1 site of collagenase is formed from the juxtaposition of three β-strands and is shaped as a deep cylindrical cavity. The negatively charged carboxylate of Asp226 extends across the base of the site and slopes downward away from substrate (Figure 1). One carboxylate oxygen of this residue accepts hydrogen bonds from the main chain amide of Thr190, as well as from a well-ordered water molecule located more deeply in the site. These interactions effectively sequester this oxygen atom from contact with a P1-substrate side chain. The other oxygen of Asp226 is oriented more in the direction

[1] Abbreviations: AMC, 7-amino-4-methylcoumarin, DMF, *N*,*N*-dimethylformamide; MUTMAC, 4-methylumbelliferyl *p*-(*N*,*N*,*N*-trimethylammonium) cinnamate; Orn, ornithine; PAGE, polyacrylamide gel electrophoresis; pNA, *p*-nitroanilide; Sbzl, benzyl thioester; SDS, sodium dodecyl sulfate; Suc-AAP-Xaa-pNA, succinyl-Ala-Ala-Pro-Xaa-pNA; Xaa, any amino acid; Z, carbobenzoxy; Z-GPR-pNA, Z-Gly-Pro-Arg-pNA.

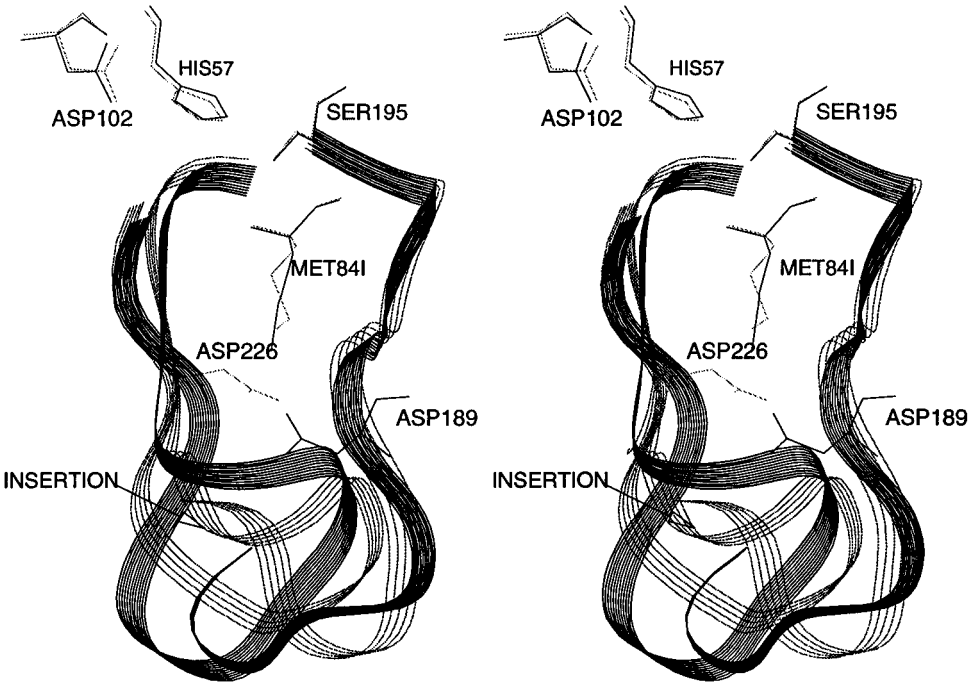

FIGURE 2: Stereoview of a superposition of the structures of fiddler crab collagenase (light ribbon) and trypsin (dark ribbon) in the area of the S1 substrate binding site. The insertion of two amino acids in collagenase, forming the enlarged S1 site, is marked and directly follows Gly216. Dark lines represent amino acids of trypsin and hatched lines those of collagenase. The alternate binding modes of the ecotin P1-Met(I)84 residue are also shown. The superposition is based on a least-squares best fit of all main chain atoms in the three catalytic amino acids Ser195, His57, and Asp102.

of substrate; however, in this complex with P1-Met84 of ecotin, interactions are made only with several waters which occupy the center of the pocket (Figure 1). The pocket is not as deep as that of trypsin since Asp226 is located closer to the catalytic residues than is Asp189 of trypsin. Its outer lip, formed from amino acids 214−220, is shaped quite differently owing to the insertion of two residues in collagenase following the conserved Gly216 (Figure 2). This provides additional pocket volume, resulting in a more elongated cavity. In each enzyme, the conserved Cys191−Cys220 disulfide bond (Figure 5a−c) is similarly positioned and likely imparts rigidity to the structure. The structures of the S1 sites and the intermolecular interactions at each of the two collagenase active sites in the collagenase−ecotin tetramer are identical within 0.4 Å, the estimated error in atomic positions.

The P1-Met84 of ecotin does not extend into the S1 pocket of collagenase. Instead, the side chain is oriented on the surface of the site (Figure 1). Hydrophobic interactions are made with the Phe215−Gly216 backbone and, on the opposite side, with Asn192 and the Cys191−Cys220 disulfide bond. This conformation is adopted presumably owing to an otherwise unfavorable juxtaposition with the charged Asp226 carboxylate. Since Met84 does not enter deeply into the pocket, there are binding sites instead for a number of water molecules, which interact with Asp226 and Thr190 as well as with the walls of the site. Presumably, binding of polar amino acids would displace many of these waters to allow direct interaction with enzyme groups. The way in which Met84 binds in this complex may be contrasted with its position bound to trypsin; in the latter enzyme, Asp189 is located deeper in the pocket, permitting the hydrophobic side chain to enter the site while still avoiding direct contact (McGrath et al., 1994; Figure 2). In both cases, the charged enzyme aspartate and the P1-Met side chain approach only within approximately 5 Å of each other. In the collagenase complex, the bent conformation of this methionine causes the sulfur atom rather than the terminal methyl group to most closely approach the negative charge (Figures 1 and 2).

In addition to the S1 site, at least one additional enzyme subsite possesses some sequence preference. By using a nucleophilic acyl transfer protocol, in which mixtures of peptides compete with water and with each other for attack on the acyl-enzyme, a significant preference of 10−100-fold was found in favor of Arg and Lys at position P1′ (Tsu et al., 1994). Hydrophobic side chains are also favored at this site, although to a lesser degree. The crystal structure shows that the S1′ site forms a shallow cavity arising from the juxtaposition of two surface loops at residues 57−62 and 34−42 (Figure 3). The Cys42−Cys58 disulfide bond and the aromatic side chain of Tyr40 form the two sides of the cavity and make van der Waals interactions with the P1′-Met85 of ecotin. These surfaces provide a rationale for the hydrophobic amino acid preference at this position. Modeling of Arg and Lys side chains at this position shows that charged hydrogen bonds could be formed by the terminal amine or guanidinium group with the backbone carbonyl oxygen atoms of His57 and Cys58. Additionally, it appears possible that the surface side chain of Asp60, which is rotated out of the S1′ site in the present structure, could reorient to form a direct or water-mediated electrostatic interaction with a P1′ basic residue of the substrate. The absence of well-defined binding clefts in the enzyme S2′ and S3′ sites is in accord with the lack of significant preference in the acyl transfer assay for the P2' and P3' residues (Tsu et al., 1994).

Five main chain hydrogen bonds are made by the amino acids at positions P1 and P3 of ecotin and are common to all serine protease complexes with protein and peptide inhibitors (Figure 1). These comprise two at the oxyanion hole main chain amides of Gly193 and Ser195, as well as

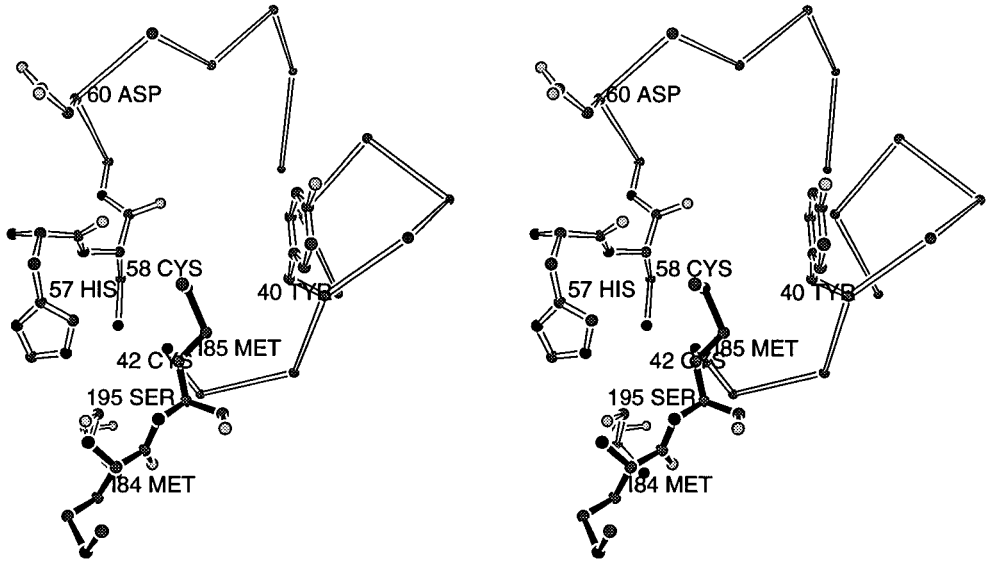

FIGURE 3: Stereoview of the S1′ binding site of fiddler crab collagenase (open bonds), shown interacting with the P1′-Met of ecotin (filled bonds). The P1-Met(I)84 of ecotin is also shown at bottom.

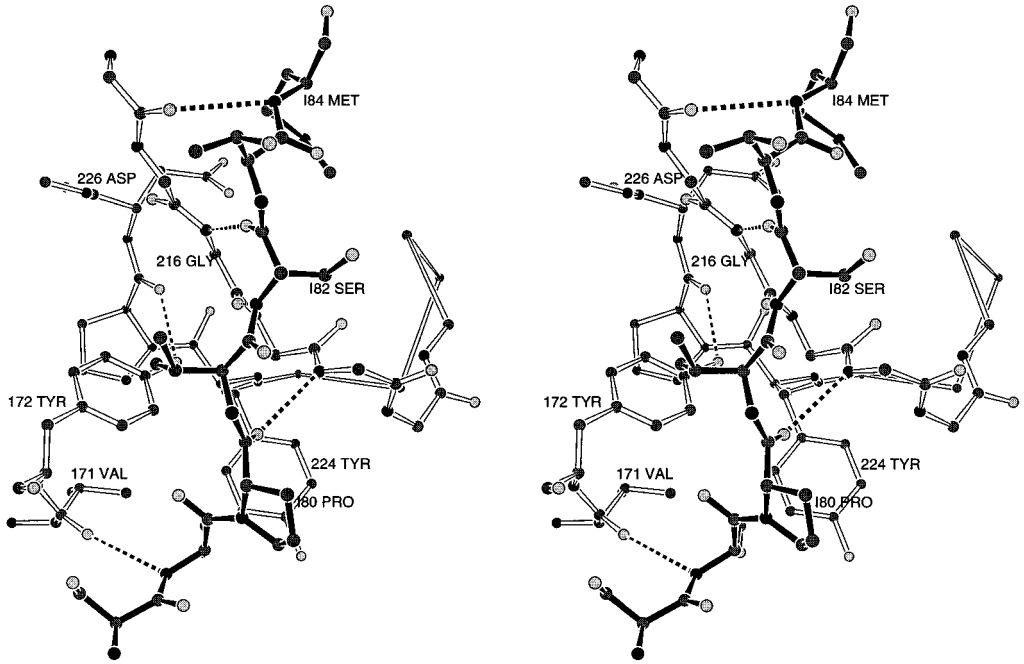

FIGURE 4: Stereoview of the binding site for the extended polypeptide chain comprising amino acid residues in the P1−P7 positions of ecotin [residues (I)78−(I)84]. Ecotin residues are drawn in filled bonds, and collagenase is drawn in open bonds. Dark dashed lines indicate hydrogen bonding interactions.

three made with the N-terminal substrate peptide at Ser214 and Gly216. Recently, the conformation of residue 216, which forms two hydrogen bonds in an antiparallel β-sheet fashion with the substrate P3 residue, has been found to correlate with P1 site preference in the mammalian but not the microbial-derived chymotrypsin-like serine proteases (Perona et al., 1995). The uniquely broad P1 site preference of the invertebrate crab collagenase does not permit its categorization as a trypsin, chymotrypsin, or elastase-like enzyme. Its backbone conformation at Gly216 is intermediate among the others and similarly does not classify it uniquely in any of the three families.

A remarkable feature of the primary interaction site in this complex is the existence of additional interactions with the P4−P7 amino acids of the substrate (Figure 4), which are mediated by several polypeptide segments of the enzyme. A small but well-defined hydrophobic pocket is present which binds the P4-Val81 of ecotin. It comprises the four amino acids Phe215, Ile99, Ile175, and Tyr172. While no kinetic data in which the substrate P4 residue is varied are yet available, the existence of the pocket predicts specificity toward smaller hydrophobic residues at this subsite. Recognition of the P5−P7 amino acids is largely mediated through two polypeptide segments at positions Ala217-Ala217(A) and Val171-Tyr172. The first of these segments directly follows Gly216 and forms a portion of the irregular loop which also shapes the P1 site. Three consecutive alanine residues followed by Gly219 form an unusual backbone conformation in which all three β-methyl groups are oriented toward the P5−P7 portion of the substrate. Snug contacts among Val171-Tyr172, Ala217, and residues Tyr224-Pro225 together form a surface upon which the P5−P7 amino

acids bind. Prominent among the intramolecular enzyme interactions is a hydrogen bond between the Tyr172 side chain hydroxyl group and the main chain carbonyl oxygen of Pro225, as well as hydrophobic contacts of the Ala217 β-methyl group with the Tyr224 and Tyr172 aromatic rings. The Val171 side chain also packs against each of the side chains of Tyr172, Tyr224, and Pro225.

The enzyme−substrate interactions at the P5−P7 sites include a number of additional hydrogen bonds involving both main chain and side chain groups of ecotin. Numerous van der Waals contacts, particularly with Val171-Tyr172, are also made. It is of interest that residue 172 has been shown to be an important determinant of the S1 site specificity of chymotrypsin, where it functions to maintain a well-ordered pocket (Hedstrom et al., 1994; Perona et al., 1995). In collagenase, the intimate contact of Val171-Tyr172 with the Ala217-Ala217(A)-Ala218-Gly219 segment similarly provides a direct means by which the more distal loop may influence the conformation of the S1 site. This structural connection could provide a pathway by which binding energy derived from interactions in the P5−P7 sites might be used in catalysis. The possible role of the extended binding site in providing collagenolytic specificity is discussed in Perona et al. (1997).

The Collagenase Specificity Profile Is Modulated by the Presence of a Negative Charge in the S1 Site. The structural gene encoding collagenase was mutated to remove the carboxylate side chain of Asp226 from the S1 site, in favor of glycine (collagenase D226G). Kinetic analysis of the mutant enzyme shows that it retains over 50% of wild-type k_{cat}/K_m versus tetrapeptide amide substrates containing hydrophobic side chains at the P1 position (Phe, Leu, Met, and Ala), but maintains only 1−5% of k_{cat}/K_m toward positively charged amino acid side chains (Arg, Lys, and Orn) (Table 1). Interestingly, the reduction in catalytic efficiency toward these three basic substrates is manifested differently in the Michaelis constants for each. The reduction in Arg activity is due to an elevated K_m, in Orn activity due to a diminished k_{cat}, and in Lys activity due to effects in both parameters. Restoration of the S1 negative charge by substituting Gly189 with Asp (collagenase D226G/G189D) recovers 40−70% of wild-type k_{cat}/K_m versus positively charged substrates, but at the cost of some additional reduction in catalytic efficiency toward hydrophobic side chains (Table 1). The k_{cat}/K_m value toward Gln is reduced by 3-fold for both variants. The sharp reduction in catalytic efficiency of collagenase D226G toward basic but not hydrophobic substrates suggests a separation in the binding sites for each of these types of amino acids. It also appears that the collagenase S1 site is quite tolerant of mutational alteration while still preserving near wild-type levels of activity, at least toward many of the preferred substrates.

Measurement of the kinetic constants of collagenase D226G and D226G/G189D toward shorter amide and ester substrates was used to assess the properties of the S1 site more directly (Table 2). The 20-fold decrease (relative to that of wild type) in k_{cat}/K_m of collagenase D226G toward a tripeptide P1-Arg substrate is very similar to an analogous decrease observed toward the P1-Arg tetrapeptide substrate (Table 1). Collagenase D226G also cleaves the single residue Z-K-Sbzl ester substrate 100-fold less efficiently than does the wild-type enzyme, whereas efficiency toward Z-Y-Sbzl is unaffected. Restoration of the negatively charged group in collagenase D226G/G189D generates an enzyme displaying wild-type levels of k_{cat} toward positively charged

Table 1: Kinetic Constants for Hydrolysis of Suc-AAP-Xaa-pNA by Wild-Type Crab Collagenase (FC) and Variants

substrate/enzyme[a] (Suc-AAP-Xaa-pNA)	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (μM^{-1} min^{-1})
arginine			
FC WT	1500	18	80
FC D226G	1400	480	2.9
FC D226G/G189D	2600	46	55
lysine			
FC WT	190	53	3.6
FC D226G	28	1100	0.024
FC D226G/G189D	310	230	1.4
ornithine			
FC WT	340	720	0.47
FC D226G	14	630	0.023
FC D226G/G189D	−	−	0.288
phenylalanine			
FC WT	2900	310	9.5
FC D226G	2350	340	6.9
FC D226G/G189D	800	580	1.4
leucine			
FC WT	1700	190	9.1
FC D226G	1300	270	4.8
FC D226G/G189D	1250	400	3.1
methionine			
FC WT	1500	360	4.2
FC D226G	1100	380	2.9
FC D226G/G189D	770	540	1.4
alanine			
FC WT	110	820	0.13
FC D226G	86	1100	0.079
FC D226G/G189D	70	1300	0.052
glutamine			
FC WT	2000	1800	1.1
FC D226G	−	−	0.33
FC D226G/G189D	−	−	0.33

[a] Reaction conditions: 50 mM Tris, 100 mM NaCl, 20 mM $CaCl_2$, and 1% DMF (or 2% Me_2SO and Suc-AAP-Gln-pNA) at pH 8.0 and 25 °C. Each value represents the mean of two determinations. Each determination included the observed hydrolysis rates for reactions at five concentrations of substrate. The standard deviations were less than 10% in all cases. A dash indicates that the individual k_{cat} and K_m values could not be determined independently due to weak binding and limited solubility of substrates. In these cases, the apparent second-order rate constant k_{cat}/K_m was determined from velocity measurements at low substrate concentrations ($[S] \ll K_m$), where $v = (k_{cat}/K_m)[E]_0[S]$ (Fersht, 1985).

Table 2: Kinetic Constants for Hydrolysis of Amide and Ester Substrates by Wild-Type Crab Collagenase and Variants

substrate/enzyme[a]	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (μM^{-1} min^{-1})
Z-GPR-pNA			
FC WT	93	230	0.43
FC D226G	2.2	100	0.022
FC D226G/G189D	220	250	0.87
Z-K-Sbzl			
FC WT	460	100	4.5
FC D226G	9.1	190	0.048
FC D226G/G189D	1400	150	9.5
Z-Y-Sbzl			
FC WT	2700	41	64
FC D226G	2600	49	53
FC D226G/G189D	625	52	12

[a] Reaction conditions: 50 mM Tris, 100 mM NaCl, 20 mM $CaCl_2$, and 1% DMF (or 9.8% Me_2SO and Z-Tyr-Sbzl) at pH 8.0 and 25 °C. Each determination included the observed hydrolysis rates for reactions at five concentrations of substrate. The standard deviations were less than 10% in all cases.

tripeptide amide and single-residue ester substrates, paralleling the results seen toward the tetrapeptide substrate. A 5-fold observed reduction in activity towards the single residue tyrosine ester also parallels analogous observations toward the tetrapeptide hydrophobic substrates. These results confirm that the observed kinetic effects seen in the battery of tetrapeptide substrates arise primarily from interactions in the S1 site, as expected. Additionally, a direct increase by a factor of 100-fold in k_{cat}/K_m upon addition of the P4 residue to the tripeptide Arg substrate is observed for the wild-type and both mutant enzymes (Tables 1 and 2). This is a substantial kinetic effect and clearly demonstrates an important role for the P4−S4 interactions in stabilizing the transition state.

Molecular Modeling of Alternate Amino Acid Side Chains at the P1 Position. A central question posed by the collagenase specificity profile is how the enzyme accommodates both hydrophobic and polar residues in the S1 site (Table 1; Tsu et al., 1994). To address this issue, we constructed models of the transition states for peptide bond cleavage of various P1 amino acids. To mimic the transition state tetrahedral geometry, the crystal structure of α-lytic protease complexed with a tetrapeptide α-aminoboronic acid inhibitor (Bone et al., 1987) was superimposed on the collagenase−ecotin structure. A best fit of the main chain atoms of the catalytic Ser195, His57, and Asp102 residues of each enzyme (rms deviation = 0.22 Å) was used. In boronate transition state analogs, the carbon of the scissile bond is replaced by boron, and the geometry is the correct tetrahedral rather than the distorted trigonal found in protein inhibitors such as ecotin. This results in a displacement of the P1 α-carbon atom by 0.75 Å in the direction away from the base of the S1 site, relative to the observed position of the ecotin P1-Met84 C^{α}. This new position was used as a starting point for the modeling experiments.

Met84 of ecotin was replaced by Gln, Arg, Leu, and Phe, and torsional angles of the mutant side chains were adjusted to fit a calculated solvent-accessible surface (Connolly, 1983) of the collagenase S1 site. Both the large P1-Phe (Figure 5a) and P1-Leu side chains fit well in approximately the same position as ecotin P1-Met84. The greater size of P1-Phe permits a greater buried surface which extends across the Asn192 backbone and the Cys191−Cys220 disulfide bond. The modeled positions of each new hydrophobic residue also necessitate a side chain reorientation of Asn192 to extend farther into solvent. In contrast, P1-Gln and P1-Arg are readily modeled to extend deeply into the S1 site where they interact with Asp226. A fully extended P1-Gln side chain could form hydrogen bonds with the side chains of both Asp226 and Thr190 (Figure 5b). Because the collagenase pocket is not as deep as that of trypsin, the larger P1-Arg side chain cannot fully extend directly downward toward the Asp226 carboxylate. However, the insertion following Gly216 creates additional volume which appears to be able to accommodate the Arg side chain in a shallower orientation. In this model, the guanidinium bridges the Asp226 carboxylate and backbone carbonyl group of Ala217, making direct interactions with each (Figure 5c). These modeling results indicate that accommodation of different amino acid side chains in distinct positions in the collagenase S1 site is in principle possible without a requirement for protein conformational change. Crystallographic analysis of variant complexes will be required to evaluate this hypothesis experimentally.

Collagenolytic Activity of the Variant Collagenase Enzymes. The collagenolytic activity of the two collagenase variants was compared directly to that of the wild-type enzyme (Figure 6). Under the experimental conditions employed, wild-type collagenase initially attacks collagen at the three-fourths cleavage site, generating the characteristic three-fourths to one-fourth fragmentation pattern (Tsu et al., 1994). This cleavage is greater than 50% complete after 2 h. Continued incubation results in cleavage of essentially all full-length substrate within 16 h, at which time secondary degradation is extensive. The one-fourth-length fragments are especially susceptible to further hydrolysis. Examination of the fragmentation patterns generated by the mutant collagenases shows that, although the rate of cleavage is slowed, the overall degradation of collagen is unchanged (Figure 6). The appearance of cleavage products in the digestion by collagenase D226G occurs in a manner parallel to that of the wild-type enzyme, albeit with an apparent 5−10-fold decrease in rate. Collagenase D226G/G189D manifests an apparent 10−20-fold decrease in the rate of collagen cleavage relative to wild-type (as judged by three-fourths fragment formation). A faint one-fourth-length triplet can be seen at 8 and 16 h. The slower rates of collagenolytic cleavage by the variant enzymes presumably arise from the same factors which cause reduced activities toward some of the short peptidyl substrates. However, an additional contributing cause to the rate reductions could be the existence of structural perturbations extending beyond the immediate area of the S1−S4 sites, which might disrupt binding interactions unique to the collagenase−collagen interface.

DISCUSSION

Specificity Profile of Fiddler Crab Collagenase. Insight into the Arg:Lys specificity ratio of crab collagenase can be obtained by comparisons of its structure with those of trypsin and trypsin D189G/G226D. Collagenase possesses an Arg:Lys specificity ratio which favors Arg by 5−8-fold, a preference very similar to that of wild-type trypsin (Table 1; Tsu et al., 1994). However, relocation of Asp189 to Asp226 in trypsin shifts the Arg:Lys specificity by 50-fold in favor of Lys, largely as a consequence of a sharp decrease of 10^3-fold in the k_{cat} toward Arg (Perona et al., 1993). This is attributable to misalignment of the scissile bond with the Ser195−His57 catalytic couple, arising as a consequence of the inability of P1-Arg to be accommodated in the shallower trypsin D189G/G226D S1 site. The relative positions of Asp226 and the catalytic groups in collagenase and trypsin D189G/G226D are similar. However, collagenase appears to be able to avoid the steric hindrance to Arg binding by virtue of its enlarged S1 site and the consequent formation of a new alternatively positioned guanidinium binding pocket (Figure 5c). The insertion of two residues following Gly216 (Figure 2) provides the basis for constructing this more extended S1 site in collagenase. The conformation of the inserted segment thus appears to be crucial to enzyme function; it is positioned to provide a binding site for P1-Arg on one side and a binding site for the extended substrate chain at positions P5−P7 on the other (Figure 4).

A more challenging goal is achieving an understanding of why hydrophobic and polar uncharged (i.e., Gln) P1

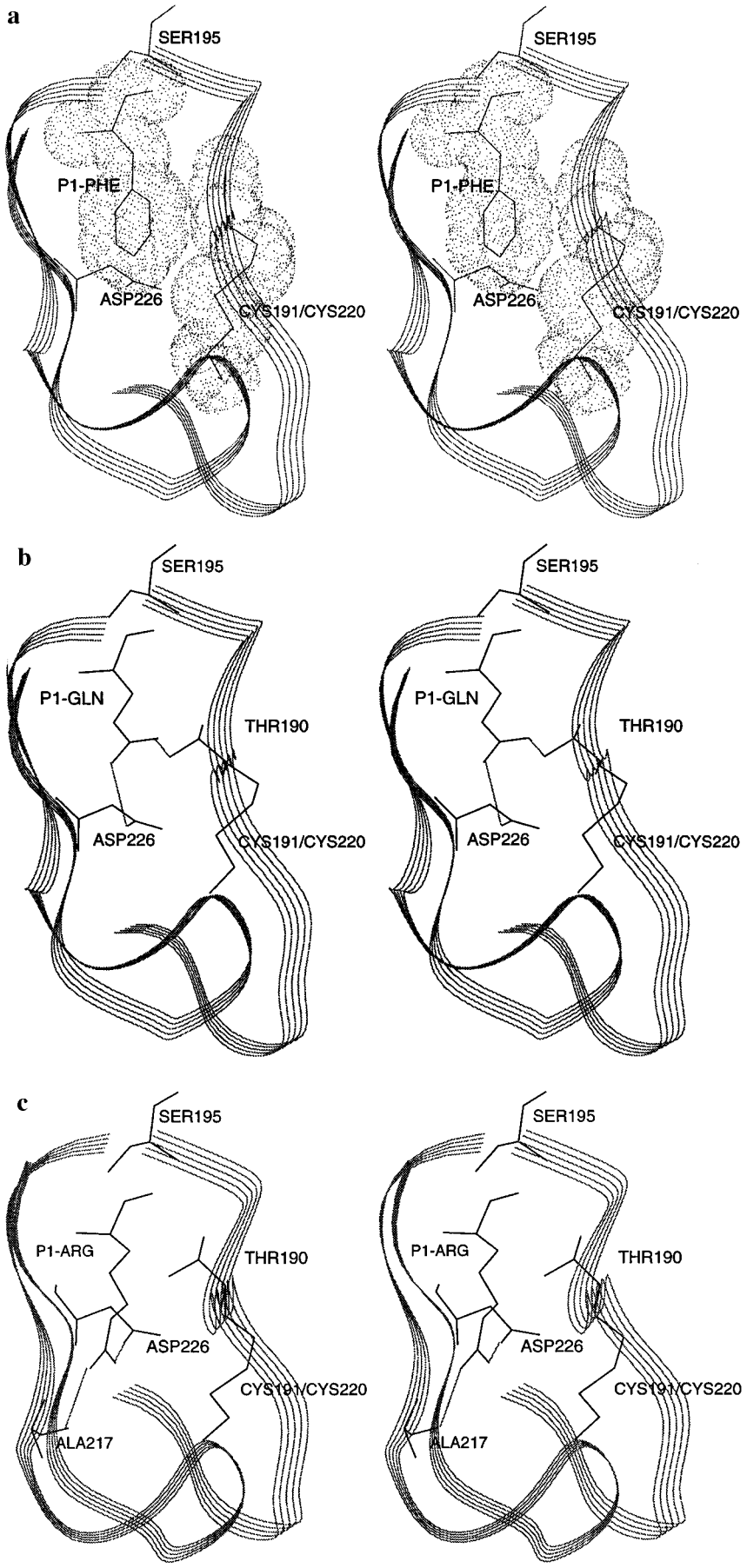

FIGURE 5: Models of the S1 site of crab collagenase complexed to (a) P1-Phe, (b) P1-Gln, and (c) P1-Arg-containing substrates. Models were constructred as described in the text. The conserved Cys191−Cys220 disulfide bond is shown at the bottom right in each figure. van der Waals surfaces of the disulfide bond and P1-Phe residue are shown in shown a. Hatched lines indicate modeled hydrogen bonding interactions of the (b) P1-Gln and (c) P1-Arg residues.

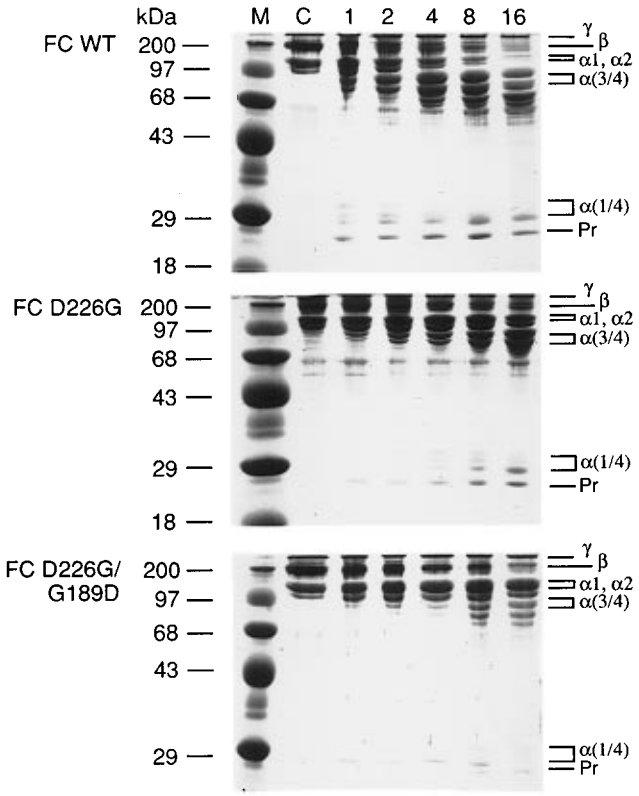

FIGURE 6: Collagen cleavage by wild-type and variant crab collagenases: lane M, markers; lane C, bovine skin collagen (primarily type I) in 50 mM Tris, 300 mM NaCl, and 20 mM $CaCl_2$ at pH 8.0; lanes 1, 2, 4, 8, and 16, collagen and enzyme (24:1 mass ratio) incubated for 1, 2, 4, 8, or 16 h at 25 °C, respectively; Pr, protease; α1 and α2, monomeric collagen chains; *β*, dimeric cross-linked collagen; *γ*, multimeric cross-linked collagen; α(3/4), three-fourths-length cleavage fragments; α(1/4), one-fourth-length cleavage fragments.

substrates are excluded in trypsin D189G/G226D (and wild-type trypsin) while being efficiently cleaved by collagenase. In serine proteases, both catalytic efficiency and P1 specificity depend on the size of the synthetic substrate. For example, both trypsin and chymotrypsin use binding of the extended peptide at positions P2−P4 to greatly enhance acylation rates of P1 cognate but not noncognate substrates (Hedstrom et al., 1992, 1994; Perona et al., 1994). In chymotrypsin, relative k_{cat}/K_m values for P1-Phe relative to P1-Lys substrates increase from just 100-fold for single-residue to 10^6-fold for extended peptide substrates. In collagenase, the role of the extended binding site is enhanced beyond even that of trypsin and chymotrypsin. The ratio of k_{cat}/K_m values for peptidyl-Arg versus P1-Arg substrates is 10^3-fold greater for collagenase than for trypsin; the analogous ratio relative to chymotrypsin for peptidyl-Phe versus P1-Phe substrates is 10-fold (Tsu & Craik, 1996). Clearly, all three of these enzymes derive an enormous amount of catalytic power by exploiting the extended peptide binding site, and the origins of their unique P1 specificity profiles do not, for the most part, arise from direct S1 site interactions. A tangible consequence of this may also be reflected in the fact that collagenase variants D226G and D226G/G189D retain collagenolytic activity; again, the extended subsite interactions appear to compensate for the loss of favorable contacts in the S1 pocket.

These kinetic data indicate that broadly specific and highly efficient amidolysis by collagenase depends on a portion of the enzyme which binds the P2-P4 substrate residues (Tables 1 and 2; Tsu & Craik, 1996). By analogy with the role of Gly216 in trypsin and chymotrypsin, we suggest that the ability of collagenase to efficiently cleave a wide range of peptidyl substrates is due to the ability of this residue to aid proper positioning of the scissile bond relative to the Ser195−His57 catalytic couple. In all trypsin-like serine proteases, Gly216 forms two main chain hydrogen bonds with the substrate P3 residue, in an antiparallel *β*-sheet fashion (Figure 1). In trypsin and chymotrypsin, the role of this amino acid as a specificity determinant has been defined; efficient acylation of cognate substrates correlates with distinct Gly216 conformations (Perona et al., 1995). Variant trypsins which do not possess chymotrypsin-like backbones at Gly216 fail to accelerate the hydrolysis of peptidyl P1-Phe amide substrates.

In collagenase, accurate substrate bond positioning apparently occurs regardless of the particular details of the distal P1−S1 interactions, which vary widely among the different P1 side chains. This represents an important distinction relative to trypsin, where the presence of Asp189 at the base of the S1 site is crucial to accurate scissile bond positioning (Perona et al., 1993, 1994). Variant trypsins in which the Asp189 carboxylate group is relocated or deleted possess wild-type conformations at Gly216, yet are severely compromised in acylation rate toward peptidyl P1-Lys/Arg substrates. Very high specificity can be achieved in this manner, because Gly216 is unable to function in substrate positioning without assistance from the ionic interactions of the P1-Lys/Arg side chains with Asp189. By contrast, we suggest that the broadly specific collagenase has acquired in the course of divergent evolution a particular Gly216 conformation which accurately positions the substrate scissile bond independently of the nature of the S1 site interactions. Perhaps not surprisingly, this conformation does not match any of the three subgroupings corresponding to trypsin, chymotrypsin, or elastase-like P1 specificities (Perona et al., 1995). In support of this hypothesis, the kinetic data for collagenase D226G/G189D show that catalytic efficiency toward Arg- and Lys-containing substrates does not depend on the position of the negatively charged carboxylate in the binding pocket (Table 1).

In collagenase, a significant additional dimension to catalysis may also be added by virtue of the extended binding site interactions. It is already clear that the P4 residue alone can contribute a kinetic effect of 100-fold relative to a substrate occupying just the P1−P3 sites (Tables 1 and 2). Of particular interest now is an exploration of possible additional kinetic effects by virtue of the unique interactions made in the S5−S7 enzyme sites. Such experiments should lead to an expanded understanding of the interrelationship between substrate binding and catalysis in the chymotrypsin-like serine protease family.

Specificity Modification in Serine and Cysteine Proteases. The S1 site of fiddler crab collagenase tolerates mutation easily, since removal or relocation of the primary Asp226 determinant does not significantly decrease activity toward many of the substrates tested. Further, significant retention of activity accompanies alteration of the substrate specificity profile. In these properties, the enzyme resembles several other serine and cysteine proteases whose behavior upon mutagenesis is similar. For example, the broad specificity of collagenase has also been observed in the S1 sites of the

bacterial serine proteases subtilisin BPN′ and α-lytic protease (Bauer et al., 1981; Bone et al., 1991), and in the S2 sites of the cysteine proteases cathepsin B and cruzain (Khouri et al., 1991; Eakin, 1992). With the exception of α-lytic protease, which possesses a specificity profile exclusively toward hydrophobic amino acids, each of these enzymes shows high levels of activity toward both hydrophobic and basic residues. The k_{cat}/K_m differential among these disparate substrates is only 10−100-fold. Removal of the negative charge thought to be important for recognition of basic amino acids in both collagenase (collagenase D226G) and subtilisin (subtilisin E156Q; Wells et al., 1987a) has remarkably similar effects on the specificity profiles of these structurally dissimilar enzymes. In subtilisin, an approximately 40−150-fold decrease in k_{cat}/K_m toward a P1-Lys-containing peptidyl amide substrate is observed upon removal of a negative charge from the S1 site, with only a small 2−5-fold effect toward the analogous P1-Met substrate. Similarly, insertion of a negative charge into the S2 pocket of cruzain in a cathepsin B-like position (cruzain V133A/S205E) causes a 20-fold increase in k_{cat}/K_m toward P2-Arg substrates, with a concomitant 5-fold drop in k_{cat}/K_m toward P2-Phe (McGrath et al., 1995). Thus, the response of this structurally and mechanistically dissimilar cysteine protease to localized mutation of a binding site also parallels that observed in collagenase.

These observations stand in sharp contrast to the requirements for specificity modification in trypsin, where 15 amino acid substitutions are required to obtain chymotrypsin-like substrate preference (Hedstrom et al., 1992, 1994). Further, attempts to convert the specificity of trypsin to that of other homologs (including collagenase) with a small number of S1 site substitutions have been uniformly unsuccessful (J. Perona, C. A. Tsu, and C. S. Craik, unpublished observations). Extensive comparisons among the structures of wild-type and mutant trypsins suggest that a primary rationale for the difficulty in modifying specificity may be a high inherent degree of rigidity in structural elements involved in transition state stabilization (Perona et al., 1995; Perona & Craik, 1995). In contrast, a similar analysis of variant α-lytic protease crystal structures suggests that an inherent flexibility of the binding pocket, particularly in the region of residue 216 and subsequent amino acids, is the crucial determinant underlying the high activities of certain variants toward new hydrophobic amino acids (Bone et al., 1989). Deformability of this part of the structure may compensate for differences in the details of the distal P1 side chain interactions in the S1 site of the variants so that proper registration of the scissile bond with the catalytic Ser195 and His57 residues can occur (Bone et al., 1989, 1991). A similar mechanism may operate in collagenase and perhaps be of relevance to other proteases in which specificity is easily modified. Establishing more definitively whether ease of specificity modification does in fact consistently correlate with an inherent capacity for structural deformability will require (along with other approaches) structural analysis of variant enzymes in many different systems. The experimental procedures developed for the mutagenesis, expression and kinetic and crystallographic characterization of collagenase variants (Tsu et al., 1994; Tsu & Craik, 1995) make this enzyme an excellent choice for the exploration of these important questions.

ACKNOWLEDGMENT

Figures 1, 3, and 4 were produced by program MAXIMAGE written by M. Rould.

REFERENCES

Bauer, C. A., Brayer, G. D., Sielecki, A. R., & James, M. N. G. (1981) *Eur. J. Biochem. 120*, 289−294.
Bone, R., Shenvi, A. B., Kettner, C. A., & Agard, D. A. (1987) *Biochemistry 26*, 7609−7614.
Bone, R., Silen, J. L., & Agard, D. A. (1989) *Nature 339*, 191−195.
Bone, R., Fujishige, A., Kettner, C. A., & Agard, D. A. (1991) *Biochemistry 30*, 10388−10398.
Connolly, M. (1983) *J. Appl. Crystallogr. 16*, 548−558.
Corey, D. R., & Craik, C. S. (1992) *J. Am. Chem. Soc. 114*, 1784−1790.
Dayringer, H., Tramontano, A., Sprang, S., and Fletterick, R. J. (1986) *J. Mol. Graphics 4*, 82−88.
Eakin, A. (1992) Ph.D. Thesis, University of California, San Francisco.
Eisen, A. Z., Henderson, K. O., Jeffrey, J. J., & Bradshaw, R. A. (1973) *Biochemistry 12*, 1814.
Fersht, A. R. (1985) *Enzyme Structure and Mechanism*, 2nd ed., p 99, W. H. Freeman, New York.
Grant, G. A., & Eisen, A. Z. (1980) *Biochemistry 19*, 6089−6095.
Grant, G. A., Henderson, K. O., & 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.
Hedstrom, L., Szilagyi, L., & Rutter, W. J. (1992) *Science 255*, 1249−1253.
Hedstrom, L., Perona, J. J., & Rutter, W. J. (1994) *Biochemistry 33*, 8757−8763.
Jameson, G. W., Roberts, D. V., Adams, R. W., Kyle, W. S. A., & Elmore, D. T. (1973) *Biochem. J. 131*, 107−117.
Khouri, H. E., Vernet, T., Menard, R., Parlati, F., Laflamme, P., Tessier, D. C., Gour-Salin, B., Thomas, D. Y., & Storer, A. C. (1991) *Biochemistry 30*, 8929−8936.
Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 488−492.
Laemmli, U. K. (1970) *Nature 227*, 680−685.
McCroskery, P. A., Wood, S. J., & Harris, E. J. (1973) *Science 182*, 70−71.
McGrath, M. E., Erpel, T., Bystroff, C., & Fletterick, R. J. (1994) *EMBO J 13*, 1502−1507.
McGrath, M. E., Eakin, A., Engel, J., McKerrow, J., Craik, C., & Fletterick, R. (1995) *J. Mol. Biol. 247*, 251−259.
Perona, J. J., & Craik, C. S. (1995) *Protein Sci. 4*, 337−360.
Perona, J. J., Tsu, C. A., McGrath, M. E., Craik, C. S., & Fletterick, R. J. (1993) *J. Mol. Biol. 230*, 934−949.
Perona, J. J., Hedstrom, L., Wagner, R., Rutter, W. J., Craik, C. S., & Fletterick, R. J. (1994) *Biochemistry 33*, 3252−3259.
Perona, J. J., Hedstrom, L., Rutter, W., & Fletterick, R. J. (1995) *Biochemistry 34*, 1489−1499.
Perona, J. J., Tsu, C. A., Craik, C. S., & Fletterick, R. J. (1997) *Biochemistry 36*, 5381−5392.
Stroud, R. (1974) *Sci. Am. 23*, 74−88.
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.
Van Wormhoudt, A., Le Chevalier, P., & Sellos, D. (1992) *Comp. Biochem. Physiol. 103B*, 675−680.
Welgus, H., Jeffrey, J., & Eisen, A. (1981) *J. Biol. Chem. 256*, 9510−9515.
Wells, J. A., Powers, D. B., Bott, R. R., Graycar, T. P., & Estell, D. A. (1987a) *Proc. Natl. Acad. Sci. U.S.A. 84*, 1219−1223.
Wells, J. A., Cunningham, B. C., Graycar, T. P., & Estell, D. A. (1987). *Proc. Natl. Acad. Sci. U.S.A. 84*, 5167−71.

BI961753U