Protein Science (1997), *6*:1806–1824. Cambridge University Press. Printed in the USA.

Crystal structures of bovine chymotrypsin and trypsin complexed to the inhibitor domain of Alzheimer's amyloid β-protein precursor (APPI) and basic pancreatic trypsin inhibitor (BPTI): Engineering of inhibitors with altered specificities

AXEL J. SCHEIDIG,*[1] THOMAS R. HYNES,*[2] LAURA A. PELLETIER,[3] JAMES A. WELLS, AND ANTHONY A. KOSSIAKOFF

Protein Engineering Department, Genentech, Inc., South San Francisco, California 94080

(RECEIVED October 28, 1996; ACCEPTED May 7, 1997)

Abstract

The crystal structures of the inhibitor domain of Alzheimer's amyloid β-protein precursor (APPI) complexed to bovine chymotrypsin (C-APPI) and trypsin (T-APPI) and basic pancreatic trypsin inhibitor (BPTI) bound to chymotrypsin (C-BPTI) have been solved and analyzed at 2.1 Å, 1.8 Å, and 2.6 Å resolution, respectively. APPI and BPTI belong to the Kunitz family of inhibitors, which is characterized by a distinctive tertiary fold with three conserved disulfide bonds. At the specificity-determining site of these inhibitors (P1), residue 15(I)[4] is an arginine in APPI and a lysine in BPTI, residue types that are counter to the chymotryptic hydrophobic specificity. In the chymotrypsin complexes, the Arg and Lys P1 side chains of the inhibitors adopt conformations that bend away from the bottom of the binding pocket to interact productively with elements of the binding pocket other than those observed for specificity-matched P1 side chains. The stereochemistry of the nucleophilic hydroxyl of Ser 195 in chymotrypsin relative to the scissile P1 bond of the inhibitors is identical to that observed for these groups in the trypsin–APPI complex, where Arg 15(I) is an optimal side chain for tryptic specificity. To further evaluate the diversity of sequences that can be accommodated by one of these inhibitors, APPI, we used phage display to randomly mutate residues 11, 13, 15, 17, and 19, which are major binding determinants. Inhibitor variants were selected that bound to either trypsin or chymotrypsin. As expected, trypsin specificity was principally directed by having a basic side chain at P1 (position 15); however, the P1 residues that were selected for chymotrypsin binding were His and Asn, rather than the expected large hydrophobic types. This can be rationalized by modeling these hydrophilic side chains to have similar H-bonding interactions to those observed in the structures of the described complexes. The specificity, or lack thereof, for the other individual subsites is discussed in the context of the "allowed" residues determined from a phage display mutagenesis selection experiment.

Keywords: Kunitz inhibitors; molecular recognition; phage display; protease–inhibitor complexes

Reprint requests to: Anthony A. Kossiakoff, Protein Engineering Department, Genentech, Inc., 460 Point San Bruno Blvd., South San Francisco, California 94080; e-mail: koss@gene.com.

*The first two authors contributed equally to the work.

[1]Present address: Max-Planck-Institut für Moleculare Physiologie, Rheinlanddamm 201, 44026 Dortmund, Germany.

[2]Present address: Central Research, Pfizer Inc, Groton, Connecticut 06340.

[3]Present address: Agouron Pharmaceutics, 3565 General Atomics Court, San Diego, California 92121.

[4]To distinguish between inhibitor and protease, amino acid numbers for the inhibitor are designated by (I). The chymotrypsin numbering system is used for the proteases.

Abbreviations: APP, amyloid β-protein precursor; APPI, amyloid β-protein precursor inhibitor domain; BPTI, bovine pancreatic trypsin inhibitor; TPCK, *N*-tosyl-/-phenylalanine chloromethyl ketone; pNA, *p*-nitroanilide; BAPA, *N*-α-benzoyl DL-arginine-*p*-nitroanilide; MCA, 4-methyl-coumaryl-7-amide; T-, trypsin; C-, chymotrypsin; wt, wild-type; RMSD, RMS deviation. Nomenclature for the substrate amino acid residues is Pn,..., P2, P1, P1′, P2′,..., Pn′, where P1-P1′ denotes the hydrolyzed bond. Sn,..., S2, S1, S1′, S2′,..., Sn′ denote the corresponding enzyme binding sites (Schechter & Berger, 1967).

A particularly fertile system for studying structure–function relationships of molecular recognition involves protease–inhibitor association. In this regard, serine protease enzymes of the trypsin-like family have received considerable attention, because they display a broad range of binding and specificity requirements and play critical roles in a myriad of important biological processes (Stroud, 1974; Neurath, 1984; Barrett & Rawlings, 1995). In many cases, the proteolytic activities are arrested by a specific inhibitor (Laskowski & Kato, 1980). The Kunitz class of inhibitors is one of a number of protease inhibitor families involved in protease regulation. These inhibitors contain about 58 amino acids and share a high degree of sequence homology (Creighton & Charles, 1987), including the distribution of three disulfide groups, and are very similar in their tertiary structures. They form 1:1 complexes with their enzyme targets by providing an extended loop that fills the enzyme active site in a conformation mimicking a bound substrate through a series of independent subsite interactions (Huber et al., 1974). With additional contacts from a second binding loop, a total of up to 15 inhibitor residues are involved in the interface, making the interaction substantially more extensive than that of a peptide substrate. Taken together, these interactions normally generate very tight complexes with binding constants in the picomolar (pM) range.

The basic principles of how molecular surfaces fit together in a complementary fashion are generally recognized. Factors that affect binding through steric, electrostatic, entropic, and solvation effects are areas of intense study; however, at this point, the details of their interplay are not understood clearly. A message emerging from the composite of considerable structural and mutagenesis work is that nature, through evolution, has established many different ways to solve its molecular recognition requirements. It is assumed that the productive complex requires a high degree of shape complementarity and that local regions within the interface must also be matched chemically in terms of polar characteristics. The interface is not a continuum, but a set of distinct subsites with different topographies and polar environments. Structural studies support this model (reviewed in Read & James, 1986; Bode & Huber, 1992).

It is noteworthy that the Kunitz inhibitors, APPI and BPTI, bind to chymotrypsin with nanomolar affinities (Wagner et al., 1992; A.J. Scheidig, unpubl. results). This is surprising because both inhibitors have a basic side chain at residue 15, the so-called P1 position. The P1 position is the side chain defining the primary specificity of the inhibitor (or substrate) and inserts directly into the enzyme's binding pocket. Chymotrypsin has a distinct preference for substrates with large hydrophobic residues at the P1 position (Baumann et al., 1970; Berezin & Martinek, 1970). The observation that there can be a mismatch at the principal binding site demonstrates that, in the presence of a number of other energetically productive interactions, one or a few such mismatches can be tolerated.

The energetic contribution of individual residues contained within the binding epitope of APPI complexed to trypsin has been assessed by alanine scanning mutagenesis (Castro & Anderson, 1996). There has also been a comprehensive analysis of the effects on affinity toward several different proteases of a variety of P1 substitutions in the context of the Ovomucoid inhibitor third domain (Lu et al., 1997). Although the interaction of the P1 residue has been shown to be the major contributor, it is expected that the relative importance of interactions at the binding subsites will vary somewhat between different protease–inhibitor pairs, and that groups of interactions might act in a synergistic manner. In fact, it has been determined that to alter trypsin specificity to that of chymotrypsin requires not only reconfiguring its binding pocket, but exchanging several distant surface loops as well (Hedstrom et al., 1992).

To evaluate thoroughly the issue of the relationship between primary and secondary specificity by standard mutagenesis techniques would require making an unrealistically large number of mutants. This problem has been overcome by employing the method of phage display mutagenesis, which allows the efficient selection of proteins that have undergone exhaustive and simultaneous mutation at a few specific sites (Wells & Lowman, 1992). Similar types of studies selecting Kunitz inhibitor mutants displaying novel binding properties have also been reported (Roberts et al., 1992; Dennis & Lazarus, 1994a, 1994b; Dennis et al., 1995; Markland et al., 1996).

The aim of the work reported here is to develop a structural perspective about the properties that define the specificity of the binding interface between these inhibitors and trypsin-like serine proteases. This study is based on the analysis and comparison of two different inhibitors, APPI and BPTI, complexed to two proteases, trypsin and chymotrypsin. It is significant that these inhibitors are targeted toward proteases with trypsin specificity; that is, they have a basic side chain in the P1 position. The presence of a basic P1 residue is counter to the specificity of chymotrypsin. How is chymotrypsin able to bind tightly an inhibitor with an arginine or lysine at P1? The structures presented here show that Lys and Arg side chains can have alternate modes of binding using different parts of the binding pocket, which explains why earlier attempts to model the chymotrypsin–BPTI complex produced an incorrect orientation of the Lys 15(I) side chain in the S1 pocket (Blow et al., 1972). The phage display mutagenesis results show the binding subsites on the enzymes can reasonably accommodate a variety of side-chain types at the five sites on APPI that were mutated. Of particular interest was the finding that His and Asn were highly selected P1 residues for chymotrypsin. This contrasts its inherent hydrophobic specificity, but can be explained based on the stereochemistry of the Arg and Lys residues at P1 seen in the inhibitor–enzyme complexes.

Results

Overall structure of the complexes

The refinement statistics and geometry of the final models are summarized in Table 1. The respective orientations of APPI and BPTI bound to trypsin and chymotrypsin are similar to those of BPTI bound to bovine trypsin and of APPI bound to rat trypsin (Perona et al., 1993a). Note however, that there are small, but real, overall orientational differences between the complexes of trypsin and chymotrypsin. These differences, which are about 3 degrees, were determined by a vectorial relationship between central atoms in the enzyme and inhibitor in each system, and are about the same magnitude as those found between the two molecules in the asymmetric unit for the chymotrypsin complexes (Table 2). No distinguishing features can be identified as the source of these small differences, and there is no obvious reason to consider them biologically important.

APPI and BPTI superimpose well with each other when compared in their complexed states; the one significant change in the main-chain conformation occurs in residues 39–41 and was de-

Table 1. *Data collection and refinement statistics for the proteinase–inhibitor complexes*

Proteinase inhibitor	Trypsin APPI	Chymotrypsin BPTI	Chymotrypsin APPI
Data collection			
Detector	FAST	FAST	Mar Plate (SSRL)
Temperature	RT	RT	100 K
Space group	C2	P 6_1	P 2_12_12
Cell dimensions (Å)	a = 94.1, b = 49.9, c = 68.8, β = 96.31°	$a = b$ = 101.6, c = 205.9	a = 70.5, b = 181.6, c = 46.3
Molecules per AU	1	2	2
Maximum resolution (Å)	1.8	2.5	2.0
Unique reflections	22,374	32,244	33,088
R_{merge} (% on I)[a]	5.9	8.7	3.0
Refinement			
Resolution range (Å)	10.0–1.8	8.0–2.6	8.0–2.1
Completeness (%)	93	89	86
Final R-factor (%)[b]	18.0[c]	19.2 (25.1)[d]	21.4 (32.4)
RMSD in			
Bond lengths (Å)	0.016	0.016	0.012
Bond angles (deg)	2.9°	1.9°	1.7°
Average B-factor (Å^2)			
for protein	33.5	35.1	36.1
No. of non-hydrogen atoms			
Protein	2,044	4,422	4,358
Water	153	158	262
Calcium[e]	1	—	—
Sulfate	—	4	—

[a] $R_{merge} = \Sigma|I - \langle I\rangle|/I$, where I = observed intensity and $\langle I\rangle$ = averaged intensity obtained from multiple observations of symmetry-related reflections.
[b] $R = \Sigma|F_o - F_c|/|F_o|$ where F_o and F_c are the observed and calculated structure factors, respectively. The *R-free* (Brünger, 1992) given in parenthesis is based on 5% of unique reflections.
[c] *R-free* was not implemented to monitor the progress of refinement.
[d] *R-free* was implemented at an advanced state of the refinement.
[e] The calcium position that has been observed in trypsin is not fully occupied in the complex or is a water molecule.

tected in the uncomplexed inhibitors as well (Hynes et al., 1990). Some differences are observed in side-chain conformations, but these are most probably a consequence of either complex formation or different crystal packing environments. The nine residues in the primary binding loop of the inhibitors (residues 11–19) superimpose well, with an RMSD in main-chain atoms of 0.25 Å, 0.23 Å, and 0.26 Å in trypsin–APPI (T-APPI), chymotrypsin–BPTI (C-BPTI), and chymotrypsin–APPI (C-APPI), respectively, using trypsin–BPTI (T-BPTI) as the reference. These differences are within the estimated error in the coordinates. The largest differences occur at the N and C termini, which are poorly ordered in all the structures.

Table 2. *Relative orientation between proteinase and inhibitor*

		Angle α (deg)	
Protease	Inhibitor	BPTI	APPI
Trypsin		165.8	167.7
Chymotrypsin	Complex 1	161.9	164.2
	Complex 2	163.3	166.8

[a] The angle is determined by two vectors, starting at C of the inhibitor residue 15 and going through the midpoint between inhibitor Cys 5-Sγ and Cys 55-Sγ, and proteinase Cys 136-Sγ and Cys 201-Sγ, respectively.

Subsites of the protease–inhibitor interface

The main chain of the principal binding loop of the Kunitz-type inhibitors (residues 11–19) forms a section of intermolecular antiparallel β-sheet structure with the enzyme binding pocket. This sheet structure consists of five main-chain H-bonds (Table 3). In this configuration, alternate residues of the inhibitor present their side chains to make productive contact with the enzyme surface. For the side-chain interactions of the Kunitz inhibitors, these are the P5, P3, P1, P2′, P4′ sites, which represent inhibitor residues 11(I), 13(I), 15(I), 17(I), 19(I), respectively.[5] A representation of this epitope and the corresponding charge and local topography on trypsin is shown in Figure 1. The S1 subsite of the enzyme and the P1 residue of the inhibitor are the principal components defining the specificity of the enzyme. A secondary inhibitor binding loop (residues 34–39) also contacts the proteases in both the trypsin and chymotrypsin complexes. Of the 14 inhibitor residues that are in contact or are proximal to the enzyme surfaces, six positions [11, 15, 17, 19, 34, and 39(I)] have been proposed as being responsible for differences in specificity (Perona & Craik, 1995). Residues in five of the six positions differ between APPI:BPTI, three on the

[5] In Schechter–Berger nomenclature (Schechter & Berger, 1967), these subsites on the enzyme have the letter designation, S(n) (for the sequential binding subsites preceding the peptide scissile bond) through S′(n) (for subsites following the cleaved bond). On the "peptide side," the corresponding letter P is used.

Table 3. *Inhibitor–protease hydrogen bonds (Å)*

	T-APPI		C-APPI	(1)		(2)	C-BPTI	(1)	(2)
Y39-Oη	S19-N	3.0							
H90-O			R17-Nη1	2.9		3.0			
F41-O	M17-N	2.9	M17-N	2.9		2.8	R17-N	2.9	3.0
L97-O							R39-Nε	3.1	3.4
D189-Oδ1	R15-Nη1	2.9							
D189-Oδ2	R15-Nη2	3.0							
S190-O	R15-Nη2	3.3	R15-Nε	3.4	R15-Nη1	3.1			
S190-Oγ	R15-Nη1	2.8							
Q192-Nε2	C14-O	2.8							
G193-N	R15-O	2.7	R15-O	2.8		2.9	K15-O	2.7	2.9
S195-N	R15-O	2.9	R15-O	2.9		2.9	K15-O	3.1	3.1
S214-O	R15-N	3.0	R15-N	3.3		3.1	K15-O	3.3	3.2
G216-N	P13-O	3.0	P13-O	3.0		2.9	P13-O	2.9	3.0
S217-O			R15-Nη2	2.9		3.1	K15-Nζ	2.8	2.8
			R15-Nη1	2.7		—			
S218-Oγ							G12-O	3.3	—
G219-O	R15-Nη1	2.8							

[a]The chymotrypsin complexes, C-APPI and C-BPTI, have two molecules in the asymmetric unit listed as (1) and (2), respectively.

first loop [15(I) (Arg:Lys), 17(I) (Met:Arg), and 19(I) (Ser:Ile)] and two on the second loop [34(I) (Phe:Val) and 39(I) (Gly:Arg)]. Structural details of the contact surfaces at these individual subsites are discussed below.

Buried surface area

On binding to chymotrypsin, APPI buries about 720 Å^2 of its surface, whereas BPTI buries slightly more, 770 Å^2. Similar numbers are found for the trypsin complexes, as well. The buried surface areas represent about 20% of their exposed surface. In both sets of complexes, there is a slightly larger percentage of hydrophobic surface buried (~55%) than hydrophilic area. As seen in Figure 2, the pattern of interactions of specific residues is quite similar, especially for the principal binding loop. However, there are some distinctive differences matching the variation in topographies and side-chain orientations at the binding interfaces of these complexes. In all cases, the major sites of interaction are through residues 15(I) and 17(I) of the inhibitors, which are highly buried and account for nearly one-half the buried surface area. A secondary site of interaction for most of these inhibitors is at position 39. Arg 39 of BPTI loses about 100 Å^2 of its accessible surface; however, this site is not important in APPI, where the residue Gly 39 loses essentially nothing upon complexation.

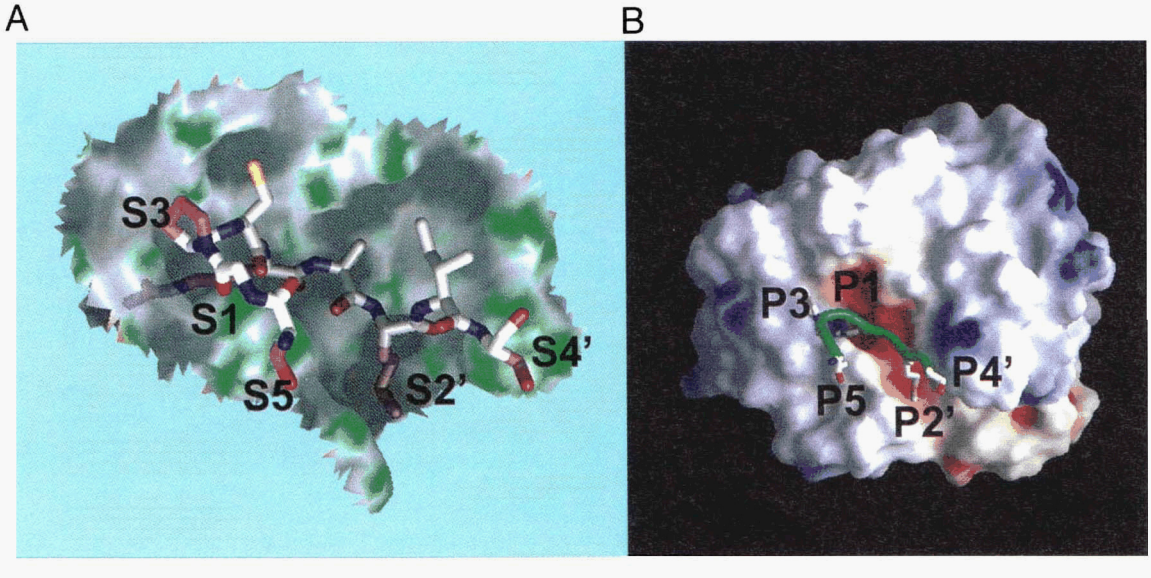

Fig. 1. A: Fitting of the APPI binding loop (residues 11–19) into the topographical surface of trypsin. Binding subsites on the trypsin surface that were characterized by phage display mutagenesis of the inhibitor are marked using Schechter and Berger nomenclature (1967). **B:** Binding loop superimposed on the electrostatic potential surface of trypsin. Inhibitor residues that were mutated are designated.

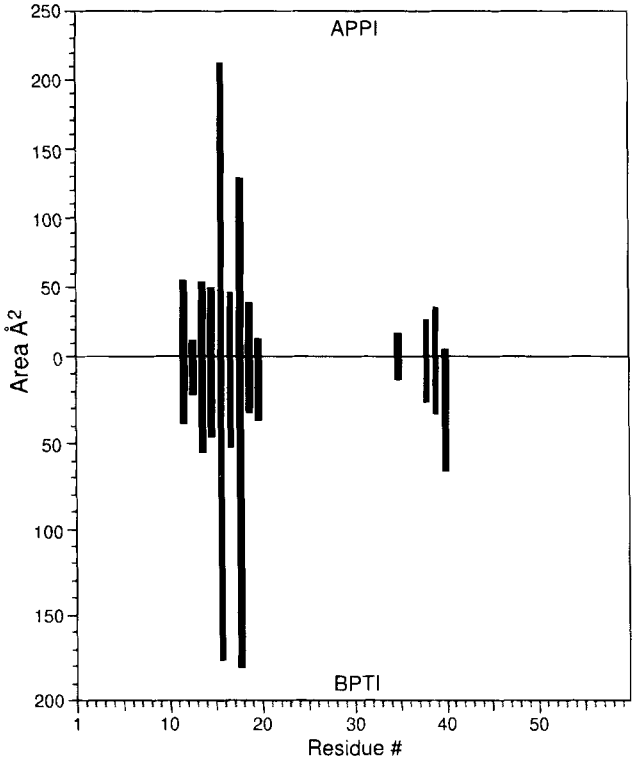

Fig. 2. Decrease in solvent accessibility on complex formation for APPI-chymotrypsin (top) and BPTI-chymotrypsin (bottom). A water probe radius of 1.4 Å was used.

S1-P1: The primary binding subsite

The S1-P1 binding site is the principal specificity-defining determinant for serine protease enzymes. As such, there are significant differences in size, shape, and hydrophilicity of the trypsin and chymotrypsin binding pockets. Because of the large variation in the residues participating at the S1-P1 interface, the details of this site are discussed separately for each enzyme.

Trypsin

The S1 binding site of trypsin is a well-defined pocket formed by residues 189–195 and 214–220. The main chain and several of the side chains of these segments are configured to make an extensive network of H-bonding interactions with the Lys and Arg P1 residues. In the T-APPI complex, the side chain of Arg 15(I) extends directly into the binding pocket, forming a direct salt bridge interaction between its guanidinium group and the carboxylate of Asp 189, which is located at the bottom of the pocket (Fig. 3A). The Arg 15(I) side chain replaces two well-ordered water molecules that fill the binding pocket of uncomplexed trypsin [Wat 414 and Wat 805 of Protein Data Bank (PDB) entry 2PTN, Marquart et al., 1983]. In addition to the salt bridge, an H-bond is made between Arg 15(I)Nη2 and the hydroxyl of Ser 190 (2.8 Å). The Nϵ of the side chain H-bonds to a water (2.9 Å), which in turn forms two long H-bonds to the carbonyl oxygens of Pro 13(I) (3.2 Å) and Gly 216 (3.3 Å).

In uncomplexed trypsin, two waters H-bond to the carboxylate of Asp 189 (Marquart et al., 1983). In the T-BPTI complex, the side chain of the inhibitor Lys 15(I) superimposes well on Arg 15(I) of its APPI counterpart. Although in T-BPTI the Lys points directly toward Asp 189, no direct interaction with the carboxylate of Asp 189 is made because of its shorter length (Huber et al., 1974). Consequently, the interaction between Lys and the carboxylate is indirect, mediated by a water molecule, which shifts by 0.8 Å toward Lys 15(I)-Nζ on inhibitor binding. An additional H-bond is made between Lys 15(I)-Nζ and a second water molecule. This conserved water position in the trypsin S1 site is anchored by H-bonds to the main-chain oxygen atoms of protease residue Val 227 and Trp 215 and to Ser 190 Oγ; in chymotrypsin, Ser 190 Oγ points away from this water. The other water molecules in the protease are not affected by the binding of BPTI.

Chymotrypsin

Although there are differences in sequence in several of the binding pocket residues and an insertion of residue 218 in chymotrypsin when compared to trypsin, the overall binding pattern of the inhibitors is essentially identical in the two enzymes. The composite differences in the binding pockets affect the stereochemical properties of the S1 sites to a significant degree. A clear difference in the binding pocket is the disposition of the side chain of residue 192 in the proteases. In T-APPI, Gln 192 is twisted into the pocket and H-bonds to Wat 5; in C-APPI, the Met 192 side chain is rotated away from the pocket and two waters (Wat 6, Wat 7) are located in the region where the Gln was found in T-APPI (Fig. 3A).

Considering the differences in topography and hydrophilicity of the binding pockets, it is not surprising that the conformations of Lys 15 (BPTI) and Arg 15 (APPI) are different between the chymotrypsin and trypsin complexes. Instead of extending directly into the pocket, as is the case for the trypsin APPI and BPTI complexes, both the Arg and Lys side chains adopt a conformation best described as a bending away from Ser 189 and toward the main chains of the inhibitor and the enzyme (Fig. 3B). This orientation is referred to as the "up" position, differentiating it from the "down" position, where the side chains extend toward the bottom of the pocket, as is seen in T-APPI and T-BPTI. In the up orientation, a set of H-bonds is made involving the main-chain oxygens of Gly 217 (2.8 Å) and Pro 13(I) (3.0 Å). The Nζ of Lys 15(I) in BPTI superimposes on Nη2 of Arg 15(I) in APPI and forms the same set of interactions (Table 3; Fig. 3B).

A comparison of the conformation of the Arg 15(I) side chain and its interactions with the binding pocket in the chymotrypsin and trypsin APPI complexes is shown in Figure 3A. The guanidinium Nϵ and Cζ move 0.7 Å and 2.1 Å, respectively, between the "up" and "down" positions. The relative organization of the water molecules is also noteworthy. A water (Wat 2) in C-APPI superimposes directly on the guanidinium Nη2 in T-APPI and Wat 3 is within 0.5 Å of the carboxylate Oδ2 of Asp 189. Conversely, Wat 5 in T-APPI is located close to Nη1 of the 15(I) side chain in C-APPI. In the C-APPI and C-BPTI structures, four water molecules assume essentially identical positions (Fig. 3B). There are, however, two additional waters in the C-APPI structure; one H-bonds to Cys 14(I) O (2.6 Å) and the other to a water that interacts with Ser 189 and Gly 221.

Based simply on stereochemical criteria, there is no apparent reason why, in the chymotrypsin complexes, the P1 side chains adopt the "up" over the "down" conformation preferentially. The Arg side chain in C-APPI can be rotated to superimpose on its T-APPI counterpart. In this position, it would displace one water molecule, Wat 2, and H-bond to two ordered water molecules, Wat 1 and Wat 3, whereas Wat 5 would be in a position to H-bond to the Nϵ, as it does in T-APPI. However, it is clear that the

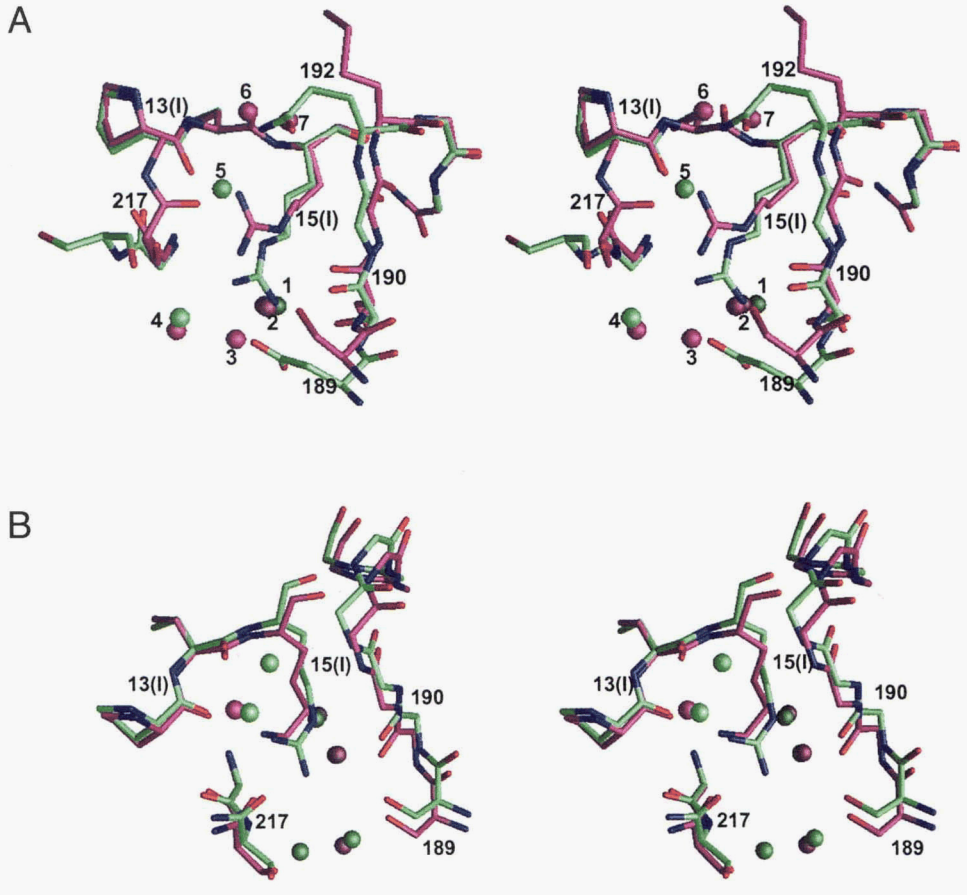

Fig. 3. A: Superposition of T-APPI (green) and C-APPI (magenta) P1 binding pockets comparing the conformations of Arg 15(I) and the local water structure. In trypsin, the Arg 15(I) side chain is in the "down" position, forming a salt bridge to Asp 189; in the chymotrypsin complex, this side chain is in the "up" position. Water molecules are color coded as above and numbered as described in the text. **B:** P1 binding pocket region of C-APPI (green) and C-BPTI (magenta) complexes. Arg and Lys side chains of inhibitor residue 15(I) are in "up" position, H-bonding to the carbonyl carbons of 13(I) of the inhibitor and 217 of the enzyme.

low-energy configuration has the Arg (or Lys in C-BPTI) side chain in the "up" position. Whatever the nature of the balance of energy that selects the "up" position versus the extended "down" position (the T-APPI conformation), it appears not to involve the shape or solvation of the chymotrypsin S1 pocket.

The S2′-P2′ subsite: An important secondary site of interaction

Composition of the S2′ binding cleft

The S2′ pocket binds the side chain of residue 17(I) (P2′) and is formed by structural components from both the protease and the inhibitor itself. The protease interface for this site is comprised of the main-chain atoms from loops 30–41 and 192–193, and by the side chains of residues 39, 40, and 151 (Fig. 4). The hydrophobic binding cleft is completed by the side chain of inhibitor residue 34(I). Residue 34(I) differs between APPI (Phe) and BPTI (Val), but is always hydrophobic.

As pictured in Figure 4, the top of the pocket is more open in chymotrypsin than in trypsin. This is partly due to the larger side chain of residue 151, which is Tyr in trypsin and Thr in chymotrypsin. Additionally, there is a proteolytic clip that removes the dipeptide, Thr 147–Asn 148, in α-chymotrypsin, the enzyme form used in this study. This clip causes some disorder in the region preceding residue 151, as well as somewhat altered conformations around 151, as seen in the two independent molecules in the crystallographic asymmetric units in both the C-APPI and C-BPTI structures. (Comparison of the two independent molecules in the chymotrypsin complexes allows for an assessment of "real" features, those observed for each structure, versus features found in only one, which are ascribed to factors such as crystal packing.) On the other hand, the Tyr 151 side chain in trypsin structure tightly encloses the 17(I) side chain.

Another prominent surface of this pocket is formed by residues 39 and 40 of the protease. Residue 40 is His in both enzymes, 39 is Tyr in trypsin and Phe in chymotrypsin, which are oriented differently in the two enzymes. In trypsin, the Tyr side chain is rotated away from 17(I) by about 60° compared to its Phe counterpart in chymotrypsin (Fig. 4). In this orientation, it makes two H-bonds: to the main-chain amide of 19(I) (3.0 Å) and to the Nζ of Lys 60 (2.9 Å) of the enzyme.

Orientations of the 17(I) side chains

The Met 17(I) side chain has a somewhat different conformation in the APPI complexes. In C-APPI, the smaller Thr 151 side chain allows the conformation of the Met side chain to be more extended compared to that in T-APPI (Fig. 4). The Cϵ carbon of the Met in

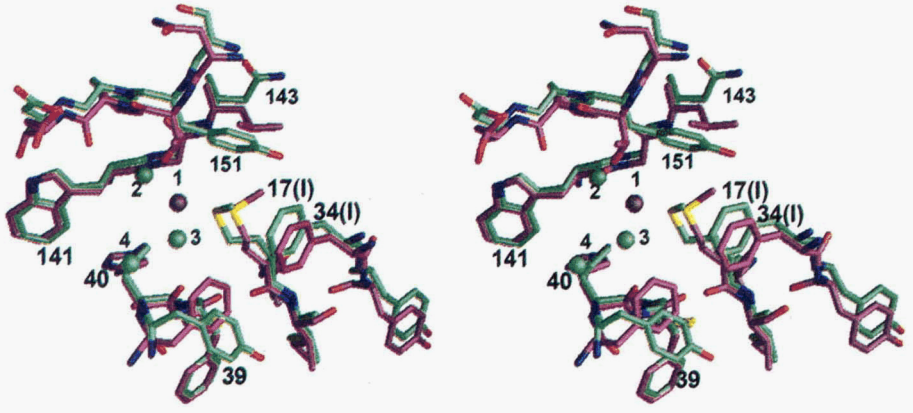

Fig. 4. Comparison of the P2′ binding pocket of C-APPI (magenta) and T-APPI (green). In C-APPI, the pocket contains one bound water molecule that H-bonds to the carbonyl oxygen of residue Trp 141 and the side chain Oγ1 of Thr 151. In T-APPI, there are three bound waters located along the edge of the site. Waters are color coded and numbered as described in the text.

trypsin almost exactly overlays Sδ in the chymotrypsin complex. Another notable difference is a translation of the Phe ring of 34(I), which keeps the ring in van der Waals contact with the Met side chain.

In each APPI complex, there are some ordered water molecules in the S2′ cleft. In C-BPTI, a water molecule (Wat 1, Fig. 4) H-bonds to the carbonyl oxygen of 141 (2.8 Å) and the Oγ1 of Thr 151 (3.3 Å). This water molecule is in van der Waals contact with the Met 17(I) Sδ. In T-APPI, three ordered water molecules are located in this pocket. Wat 2 makes an H-bond to 141 O (2.7 Å) analogous to Wat 1 in the chymotrypsin complex, but it is displaced by about 2.5 Å because of the altered orientation of the Met side chain. Wat 4 H-bonds to the main-chain amide of His 40 (2.9 Å) and to Wat 3, which, in turn, forms a long H-bond to the carbonyl oxygen of His 40 (3.3 Å).

In the BPTI complexes, Arg 17(I) packs its methylene side-chain carbons into the hydrophobic cleft described above with its hydrophilic guanidinium group extended out to make H-bonding contacts with surface polar groups and solvent. In both complexes, there is an H-bond between the carbonyl oxygen of His 40 and the Nη1 of 17(I). Additionally, there are several ordered waters that interact with the exposed guanidinium group.

The S3-P3 subsite

The sequence of residues 12–14 is Gly-Pro-Cys in both APPI and BPTI, and these residues adopt the same conformation in all the complexes. There is a strong requirement for a Gly at position 12, because there is no space for a side chain due to the proximity of the main chain of Cys 38(I), which in turn is fixed in its orientation by the disulfide bridge with Cys 14(I). Although a proline at residue 13 is reasonably well conserved in the Kunitz family of inhibitors (Creighton & Charles, 1987; Creighton & Darby, 1989), from a structural standpoint it appears that it could be replaced by a variety of amino acid types because its β carbon is oriented toward a solvent-exposed region that includes a few well-ordered water binding sites.

Subsites S5-P5 and S4′-P4′: The edge of the binding pocket

The S5 and S4′ sites define the extremities of the principal protease–inhibitor binding surface. A notable difference between the two sites is that, whereas the S4′ sites of trypsin and chymotrypsin have generally similar topographies, their S5 sites are quite different, as pictured in Figure 5. The S5 pocket is formed by residues 146, 192, and 217–219 of the enzyme and 34(I)–35(I) of the inhibitor. The posttranslational excision of the dipeptide Thr 147–Asn 148 in α-chymotrypsin, combined with sequence variations, results in a different packing of side chains 143–146 and 149–151 compared to the 143–151 loop in trypsin. In trypsin, this loop assumes a conformation where no residues are close to the Thr 11(I) side chain. Additionally, the disposition of enzyme residues 216–219 is quite different, partly due to the insertion of Thr 218 in the chymotrypsin sequence. This segment of chain in chymotrypsin is translated about 3 Å toward 11(I), where it forms an extensive interface burying about 70% of the surface area of the Thr(I) side chain. In contrast, only about 25% of this side chain becomes buried in the trypsin complexes. The surface of this side of the pocket is extended in the chymotrypsin complexes by the side chain of Tyr 146, which packs against Thr 218. The main chain of 146 in trypsin is 3 Å farther away; in addition, the side chain points away from the pocket, resulting in an open pocket in the trypsin complexes.

Another extensive face of the S5 pocket is formed by the main chain of inhibitor residues 34(I)–35(I) and the protease side chain of 192. Although the type and conformation of the 192 side chain differs between the two enzymes, the surface buried upon complexation is similar (= 85 Å^2). The chymotrypsin complexes have two ordered waters that superimpose the terminal Oε2 and Nε1 atoms of the Gln 192 side chain, as it is oriented in the trypsin complexes. A characteristic of this pocket in both the trypsin and chymotrypsin complexes is a highly organized water structure. A few waters are conserved in both structures, but most are not (Fig. 5). In all the complexes, there is room to accommodate a large P5 side chain, although this would eliminate the H-bond in the complexes between the Thr 11(I) Oγ1 and the carbonyl of Phe 34(I) (2.9 Å).

The S4′ site fits the 19(I) side chain and is formed principally by side chains of the inhibitor: 21(I) (Trp:Tyr, APPI versus BPTI), 32(I) (Pro:Thr), and 34(I)(Phe:Val) and the protease side chain 39 (Tyr:Phe, for trypsin versus chymotrypsin). In the T-APPI complex, this binding cleft is more highly structured due to the bulkier Phe 34(I) side chain and the conformation of the Tyr ring of residue 39 (Fig. 6). In this complex, Ser 19(I) H-bonds to a water molecule (Wat 1, 2.8 Å), which sits between the side chains of

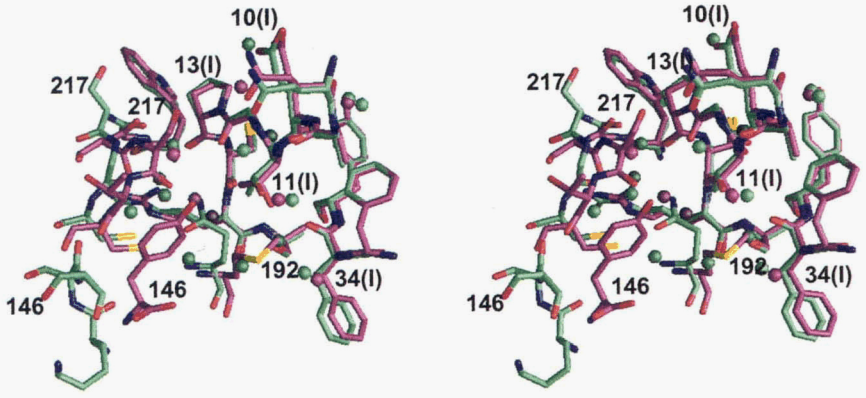

Fig. 5. Comparison of the P5 binding pocket of C-APPI (magenta) and T-APPI (green). This pocket is considerably smaller in the chymotrypsin structures than in the trypsin counterpart. Note the pattern of the ordered waters in T-APPI (green), which lay in close proximity to the main chain of 216–218 in C-APPI.

32(I) and 34(I). Although an H-bond between the hydroxyls of Ser 19(I) and Tyr 39 can be modeled by rotating the 19(I) side chain, the geometry of the resulting H-bond is not optimal and it is apparent why the energetics favor the observed configuration. Note, however, that small adjustments can make the Ser 19(I) Oγ–Tyr 39Oη interaction favorable as observed in the APPI–rat trypsin structure (Perona et al., 1993a).

In the BPTI complexes, the Ile 19(I) side chain buries about 40% of its accessible surface area. It is surrounded by a lattice of water molecules just outside its van der Waals contact surface. One of the waters in C-BPTI overlays the hydroxyl of Tyr 39 directly in the trypsin complexes and makes the same H-bond to the amide peptide nitrogen of 19(I) (3.0 Å).

Selection of APPI inhibitors with altered specificity by phage display mutagenesis

Library selection

To further elaborate on the variety of inhibitors that can bind tightly to trypsin and chymotrypsin, we randomly mutated APPI at positions 11, 13, 15, 17, and 19 using phage display and selected those mutants that bound tightly to each protease (Fig. 7). The number of DNA sequences possible when five codons are mutated randomly by substitution with NNS is 3.4×10^7 (32^5). The size of the library screened (9×10^6 independent transformants) was less than that needed to cover all possible DNA sequences, but should have provided a reasonable sampling of the 3.2×10^6 possible amino acid combinations. To estimate the sequence bias in the starting library, 27 independent clones were picked at random and sequenced. From the 135 NNS codons sequenced (5×27), the nucleotide frequencies were 19% A, 28% T, 23% C, and 30% G for the first two positions, and 48% C, 51% G, and 1% T in the third position. The starting library contained an even distribution of amino acids in each position, dependent primarily on the theoretical number of codons for each amino acid.

Libraries were selected for binding to each of the two proteases [T^+ or C^+], either in the presence [T^- or C^-] or absence [T^0 or C^0] of the other protease in solution. This produced four different selection protocols abbreviated: [T^+ C^0], [T^+ C^-], [T^0 C^+], or [T^- C^+]. Following 6 or 12 rounds of binding enrichment, the sequence diversity in the starting library was reduced to distinct distributions of amino acids (Table 4; Fig. 8). The degree of selectivity at each randomized inhibitor position was protease dependent. Positions 11 and 17, which were diverse in the optimal trypsin library [T^+ C^0], were focused more narrowly in the optimal chymotrypsin library [T^0 C^+]. Position 13 was focused in the optimal trypsin library [T^+ C^0] and diverse in the optimal chymotrypsin library [T^0 C^+].

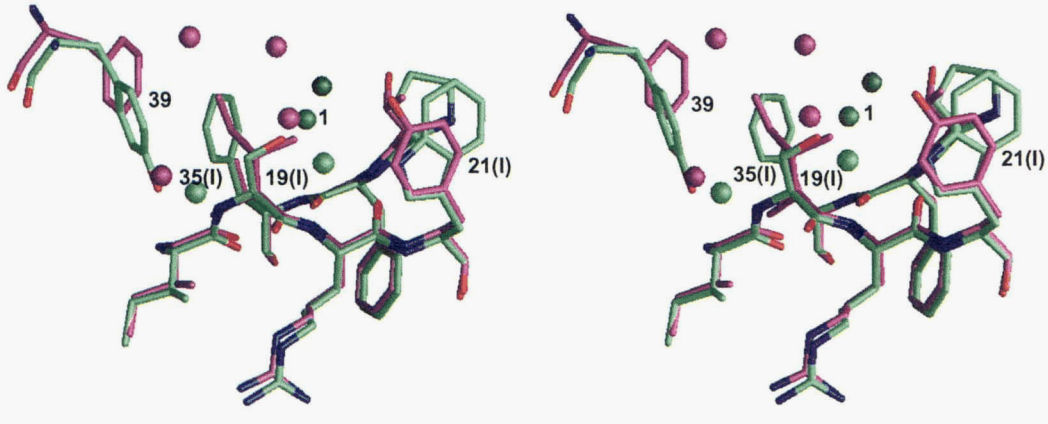

Fig. 6. S4′ binding site with overlaid structures of T-APPI (green) and C-BPTI (magenta).

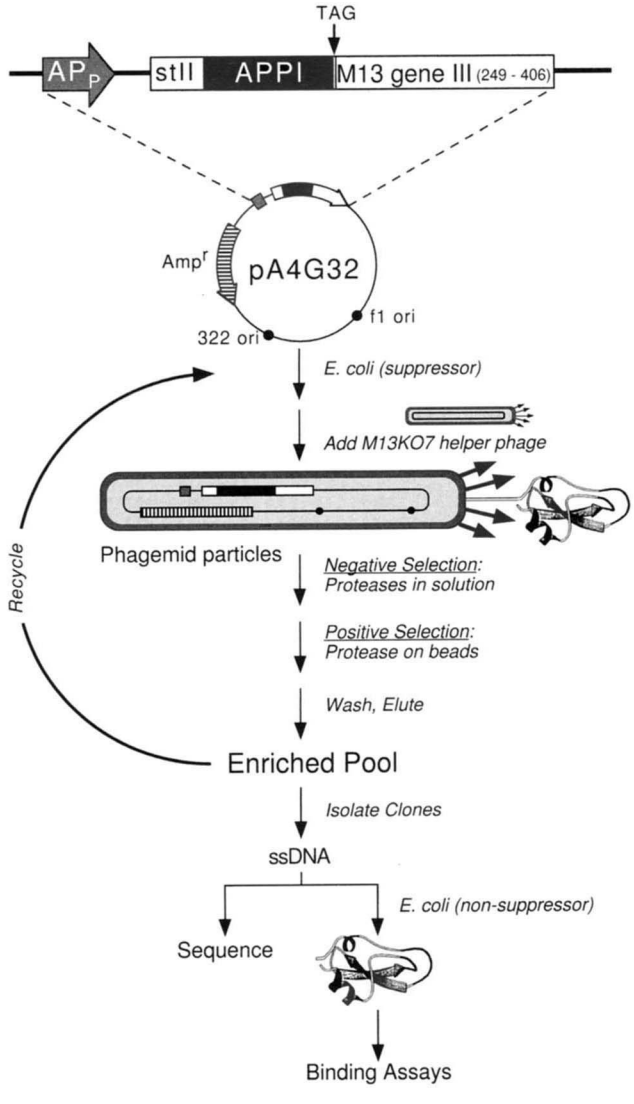

Fig. 7. Outline of the phage display vector and the strategy for positive and negative selection of APPI variants. The plasmid pA4G32 contains the alkaline phosphatase promoter, stII secretion signal, the APPI gene, and codons 249–406 of the M13 gene III protein. The starting library of phagemid particles is subjected to rounds of selection to isolate an enriched pool of sequences. Positive selection is to a protease immobilized on beads. Negative selection is implemented by pre-incubating with proteases free in solution. Broadly specific sequences are selected through successive rounds of positive selection to each protease target. After binding selection, individual phage are isolated and the ssDNA sequenced to determine the amino acid substitutions present. Finally, the ssDNA is transformed into a non-suppressor *E. coli* strain, in which the amber codon terminates translation and free inhibitor is produced.

Optimal binding sequences

In the optimal trypsin library [T^+ C^0], Arg and Lys were the only amino acids selected in position 15, the P1 position. The strong selection for Arg and Lys residues matches the substrate specificity as well as the general presence of basic residues in the P1 position of protein inhibitors of trypsin. Both APPI and BPTI bind extremely tightly to trypsin using Arg and Lys residues in position 15, respectively. The preference for Arg over Lys correlates with the higher turn-over of Arg-containing small peptide substrates relative to Lys substrates (Perona et al., 1993b). In addition, there were strong preferences for His and Asn at position 13 and Lys and Gln at position 19.

The striking result from the optimal chymotrypsin binding library [T^0 C^+] is the strong selection for Asn and His residues in position 15. The cleavage specificity of chymotrypsin is biased toward large- and medium-sized hydrophobic side chains (Bauer et al., 1976). A single Phe sequence was isolated at position 15 among the clones sequenced. This clone, as well as clones with an Asn or His in position 15, were expressed and assayed for chymotrypsin binding activity (Table 5). All three clones inhibit chymotrypsin in the low nanomolar range. This argues that, in the context of the APPI inhibitor, scaffold Asn and His side chains in the P1 site are equally potent for chymotrypsin inhibition as a large hydrophobic side chain. In position 11, Arg and His were selected with a significantly narrowed distribution after 12 rounds, compared with 6 rounds, of binding selection. Phe and Lys were preferred at positions 17 and 19, respectively.

Protease inhibition by selected sequences

The population of sequences from each selected library was expressed as a pool and assayed for inhibition of trypsin and chymotrypsin (Table 6). Values for K_i of less than 20 pM could not be determined accurately with the assay used. Individual clones from the selected libraries were expressed and assayed to determine if the apparent K_i values of the pools were representative of individual sequences (Table 5). The clones chosen for assays were sequences that most closely matched the consensus sequence of the selected libraries.

Two libraries were selected for binding to trypsin with increasing use of negative selection against chymotrypsin (Table 6). The optimal trypsin library [T^+ C^0] had a low pM K_i for trypsin inhibition, but cross-reacted with chymotrypsin with a K_i value in the 10 nM range. The addition of negative selection to chymotrypsin [T^+ C^-] decreased the cross reaction with chymotrypsin by 60-fold. However, the increased specificity of the negatively selected library, [T^+ C^-], was linked to a fourfold loss of affinity for trypsin. Thus, the use of negative selection increased significantly the specificity of the selected inhibitor sequences for their target proteases accompanied by some loss of absolute binding affinity.

Similar trends in binding specificity were seen for the libraries selected for chymotrypsin inhibition (Table 6). The K_i for chymotrypsin with simple positive selection [T^0 C^+] was in the 10 nM range and increased in the negatively selected library, [T^- C^+]. In the absence of selection (0), the K_i values for trypsin were greater than 200 nM. This represents the extent of nonspecific cross reaction that occurs in the absence of negative selection. K_i values for chymotrypsin ranged from 10 to 20 nM. This is surprising, given the fact that the library that was positively selected for chymotrypsin binding exhibited K_i values in the same range. It appears that the sequences selected for trypsin binding interact with chymotrypsin as well as sequences selected for chymotrypsin binding.

Discussion

Exploring specificity and its structural basis using phage display mutagenesis

A pertinent characteristic of the Kunitz family of inhibitors is that they bind extremely tightly to a broad range of serine proteases

Table 4. *Sequences of random clones from selected libraries*

	Position						Position				
Round 6	11	13	15	17	19	Round 6	11	13	15	17	19
$[T^+ C^0]$						$[T^0 C^+]$					
1	T	H	R	R	K	1[a]	R	R	R	F	E
2	N	Y	R	A	K	2[a]	R	R	R	F	E
3	S	S	R	K	R	3	G	H	N	F	K
4	T	N	K	T	R	4	R	K	Q	T	K
5	S	H	R	K	S	5	T	H	N	H	K
6	K	H	R	T	K	6	H	R	R	N	Q
7	S	T	R	K	T	7	P	L	R	S	L
8	I	D	R	N	Q	8	R	H	H	Y	T
9	S	L	R	R	E	9	S	R	R	S	T
						10	P	R	N	F	L
						11	P	R	H	S	T
						12	R	N	H	W	P
						13	R	H	N	H	T
						14	R	N	N	N	K
						15	R	N	N	I	E
						16	T	Q	H	K	P
						17	P	R	N	A	T
						18	R	D	Q	T	K
						19	K	S	H	H	P
						20	S	S	R	F	L
						21	R	H	N	F	P

	Position						Position				
Round 12	11	13	15	17	19	Round 12	11	13	15	17	19
$[T^+ C^0]$						$[T^0 C^+]$					
1	R	L	R	A	L	1[a]	R	Y	H	N	A
2	S	H	R	S	Q	2[a]	R	Y	H	N	A
3	P	T	R	R	E	3	H	K	R	M	P
4	R	N	R	H	R	4	R	T	N	F	K
5	P	N	R	S	T	5	R	E	N	F	K
6	T	H	R	R	Q	6	R	K	H	F	K
7	S	N	R	M	K	7	R	S	N	I	K
8	P	R	K	T	R	8	R	A	H	F	K
9	T	N	R	V	K	9	H	R	F	M	E
10	T	H	R	H	L	10	R	S	H	F	K
11	R	N	R	A	E	11	R	R	N	I	Q
12	S	N	R	E	Q	12	R	S	H	Y	K
13	S	C	R	R	Q	13	R	M	H	F	T
14	T	N	R	T	K	14	R	W	H	F	P
15	T	N	R	R	Q	15	T	R	N	F	K
16	S	S	R	T	R	16	R	H	N	F	P
17	S	T	R	A	Q	17	R	Y	N	F	K
18	Q	R	R	N	K	18	H	R	N	M	E
19	T	H	R	H	K	19	R	H	N	I	E
						20	R	Y	N	F	K
						21	R	N	N	F	E

[a]Indicates sequences that were identical at the DNA level to adjacent sequence.

whose scaffolds defining their binding surfaces are formed by quite different side-chain types. The clear implication is that some elements of the interface play more dominant roles than others. Although the P1 side chain remains the most important effector to tight binding, these inhibitors have a surprisingly broad specificity profile as demonstrated with the structures described here.

There is an extensive literature dealing with sequence specificity for the trypsin-like serine proteases (Graf et al., 1982; Evnin et al., 1990; Keil, 1992). The compilation of the data shows clearly that trypsin has a very strong preference for basic side chains at P1, whereas chymotrypsin shows a distinct preference for large hydrophobics like Tyr, Trp, and Phe. It has also been recognized that, although the extent may differ from system to system, secondary specificity plays an important role in catalysis (Hedstrom et al., 1994; Perona & Craik, 1995). In the serine protease family, the data suggest that the influence of these other subsites correlates with how optimally the P1 residue interacts with its binding pocket (Keil, 1992). For instance, in the case of chymotrypsin, Phe (or

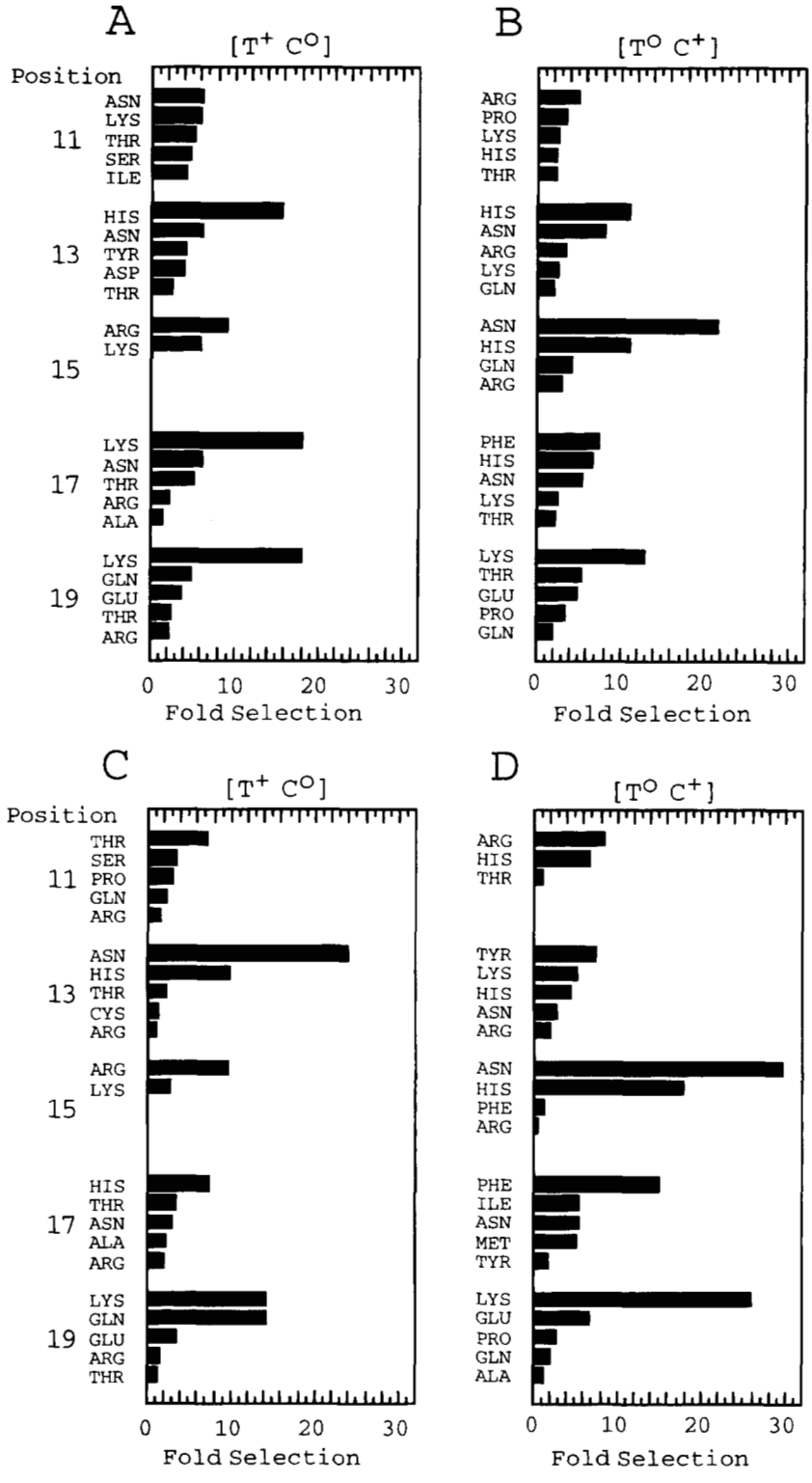

Fig. 8. Histograms of the enrichments of amino acids in the five randomized positions for libraries selected for optimal binding interaction to trypsin and chymotrypsin. Sequences following 6 rounds of selection for trypsin [T^+ C^0] and chymotrypsin [T^0 C^+] (top) and after 12 rounds of selection (bottom) are shown. For the five positions randomized (11, 13, 15, 17, 19), the top five amino acids found in each position are plotted. The fold enrichment is the observed frequency of each amino acid in the selected library divided by the calculated frequency of that amino acid in the starting library.

Table 5. *Inhibition constants for individual clones from the selected libraries*[a]

Library selection	Sequence P_5 11	P_3 13	P_1 15	P_2' 17	P_4' 19	K_i (nM) Trypsin	K_i (nM) Chymotrypsin
$[T^+ C^0]$	T	N	R	R	Q	<0.02	9
$[T^+ C^0]$	T	H	R	H	K	0.05	33
$[T^0 C^+]$	R	T	N	F	K	1,500	39
$[T^0 C^+]$	R	S	H	F	K	1,200	23
$[T^0 C^+]$	H	R	F	M	E	2,900	24
APPI[b]	T	P	R	M	S	<0.02	7.1
BPTI[b]	T	P	K	R	I	<0.02	16

[a]Assay conditions are described in Materials and methods.
[b]Purified protein was used in assays of APPI and BPTI inhibition.

Tyr, Trp) at P1 dominates substrate binding and turnover and there is only a small influence of residues at P1′ and P2′. For the weaker P1 residue, Leu, the amino acid types at P3, P2, P1′, and P2′ play a measurable role in binding. This trend continues with other less optimal P1 groups. Interestingly, for this enzyme, there is a statistically significant preference for Arg and Lys at P1′, whereas the converse is true in trypsin, where these residues have a negative influence. In this regard, it should be pointed out that there are certain subsite sequences that have a strong negative influence on substrate turnover for all these enzymes. For example, Pro at P1′ is essentially totally disallowed and Asp at either P1′ or P2 is highly disfavored.

Phage display mutagenesis was used to establish whether similar trends exist for these Kunitz inhibitors and to explore issues involving binding energies and specificity. In the procedure, five residues in the principal binding loop of APPI were targeted for random mutagenesis: positions 11, 13, 15, 17, and 19. In the context of the inhibitor structure, the binding loop does not have the conformational flexibility that a free substrate would possess. The phage mutagenesis results showed, as expected, that trypsin selections at P1 are limited to Arg and Lys and this takes preference over any secondary specificity. This is consistent with the trend that establishes dominance for the good P1 residues over secondary specificities.[6] The picture for P1 selectivity is quite different for chymotrypsin. Two features stand out. First, inhibitors with the expected P1 residues were not isolated in the selection, and second, this apparently led to greater influence of the subsites. Of 21 sequenced clones, only 1 had a Phe at P1, 11 clones had an Asn, and 8 others had a His; these latter two residue types require a hydrophilic environment. It is important to note that there is no obvious technical reason that Phe was prejudiced against, because a Phe showed a selective preference assayed against a mixture of proteases in a parallel sort using the same libraries (data not shown). The structure determination of the Asn and His P1 mutants is underway and preliminary interpretation of the difference maps indicates that these side chains form H-bonds to the edge of the binding pocket (A.J. Scheidig & A.A. Kossiakoff, unpubl. results).

[6]Inhibitors from other classes, for instance, ecotin and a-antitrypsin, having a Met at their P1 position, have more extensive interactions with nonbinding pocket residues, which are the driving force for the developed binding energy.

Table 6. *Apparent inhibition constants for selected libraries expressed as pools*[a]

Selection	Number of rounds	K_i (nM) Trypsin	K_i (nM) Chymotrysin
$[T^+ C^0]$	12	0.03	16
$[T^0 C^+]$	12	220	7.5
$[T^+ C^-]$	6	0.10	1,000
$[T^- C^+]$	6	2,000	18

[a]See Materials and methods for selection and assay details.

The chymotrypsin P1 selections are not surprising in light of the alanine-scan data reported by Castro and Anderson (1996). Those data showed that the K15(I)A mutation results in a decrease in binding of only about threefold in C-BPTI. This is in contrast to what is seen in the OMTKY3 protease–inhibitor complexes, where an alanine substitution at P1 (position Leu 18 in OMTKY3) reduces its binding affinity to chymotrypsin by almost 10^4 (Lu et al., 1997). Part of this discrepancy is due to the significantly different affinities the two classes of inhibitors inherently have to chymotrypsin. Although APPI and BPTI bind tightly to chymotrypsin ($K_i = 20 \times 10^{-9}$ M), they probably have evolved to inhibit preferentially enzymes with trypsin specificity ($K_i \sim 10^{-14}$ M). The opposite is probably true for OMTKY3, which binds to chymotrypsin with a K_i of about 2×10^{-11} M (Lu et al., 1997). This suggests that each inhibitor family has its own special set of composite stereochemical elements to bind effectively to their cognate enzyme targets.

Topographical character of the subsites

Of the five binding sites probed by phage mutagenesis, only two, S1 and S2′, have the character of a real pocket. The other three sites are formed from relatively shallow surface clefts. The pockets do not form highly restrictive environments, as can be judged by the wide variety of side-chain types that are accommodated (Table 4). A characteristic of subsites S5, S3, S2′, and S4′ is that part of the structure of the site is defined by proximal residues in the inhibitor. That is, with the exception of S1, none of the sites are preformed prior to inhibitor binding, the inhibitor itself "bringing along" a part of the pocket.

Space-filling renderings of the S5 (position 11) and S2′ (position 17) subsites in T-APPI are shown in Figure 9. In the phage sorts for C-APPI, the P5 residue, position 11, had a predilection for Arg and there was a strong preference for a large hydrophobic group at P2′, position 17. A charged side chain is preferred at position 19. Preliminary structure analyses of several mutants indicate that accurate modeling of the side chains into the subsites would be extremely difficult, because the orientation of the side chains of either the inhibitor, the enzyme, or both can change, as well as the interacting water structure (A.J. Scheidig & A.A. Kossiakoff, unpubl. results).

The S5 subsite

Thr 11(I) lies in a shallow cleft at the edge of the binding pocket (Fig. 9A). The orientation of the side chain is fixed through an H-bond to Phe 34(I) O. The favored residues in the trypsin phage sorts are Ser and the wild-type residue Thr. Presumably, the same H-bonding interaction takes place in the serine mutants. In chy-

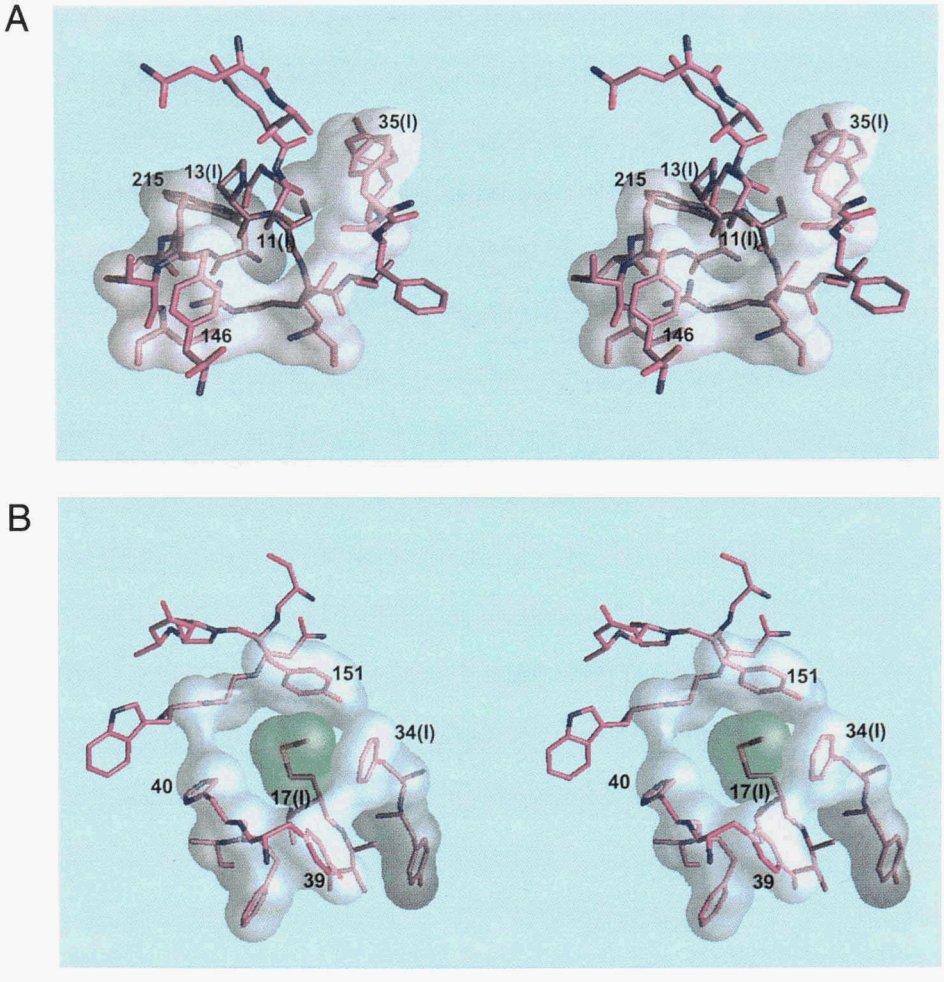

Fig. 9. Space-filling renderings of subsite binding pockets S5 (**A**) and S2′ (**B**) in T-APPI.

motrypsin, there is a strong preference for Arg; this large side chain can be accommodated easily by its extension into the accessible solvent region. An interesting point in this regard is that, although the disposition of the Arg side chain in chymotrypsin is restricted to a more confined binding cleft, in trypsin several possible orientations for a basic side chain exist. In a T11K T-APPI mutant, the Lys side chain is oriented to H-bond to the Oϵ2 of Gln 192 (unpubl. results). This conformation is not available to mutants binding to chymotrypsin not only because residue 192 is a Met, but also because the ring of Tyr 146 and the main-chain 217–218 would interfere (Fig. 9A).

The S2′ subsite

Along with S1, the S2′ subsite is a major contributor to binding based on its ability to bury large side chains of either hydrophobic or hydrophilic character. In APPI and BPTI, the 17(I) side chain contributes about 140 Å^2 of buried surface area to the complex. As discussed above, the pocket is somewhat smaller in trypsin due to the "capping" by the Tyr 151 side chain (Fig. 9B). Given these similarities, it is noteworthy that the chymotrypsin selections are almost exclusively large hydrophobics, whereas those for trypsin are generally hydrophilic and a significant percentage of them are small. Surprisingly, Ala was found in 3 of 19 clones in the trypsin sorts. This produces a significant hydrophobic cavity. It is probably the case that the energetics of binding are so highly driven by the P1 residue that there is little dynamic range in the binding energies associated with the S2′ pocket.

Stereochemistry at the attacking nucleophile

The stereochemistry around the catalytic machinery in structures of Kunitz inhibitor–protease complexes is thought to model the structure of a Michaelis complex. It has been proposed that a principal determinant of efficient enzyme hydrolysis involves the alignment of the attacking nucleophile to the scissile bond of the substrate (Kossiakoff, 1987). The complementarity of the P1 side chain to the enzyme binding pocket presumably plays a large role in the precise orientation of the scissile bond. Judged from the P1 binding observed in the trypsin complexes, the configurations of the catalytic Ser 195 and the P1 scissile bond represent productive stereochemistries, whereas those of the chymotrypsin complexes might be affected by the less than optimal binding of the Arg (APPI) and Lys (BPTI) side chains. Thus, comparison of the trypsin and chymotrypsin complex structures can provide some insight into the degree to which binding of nonoptimal P1 residues might distort the stereochemistry of the attacking nucleophile at the scissile bond.

To make such comparisons, it is important to eliminate potential artifacts stemming from stereochemical restraints applied during

structure refinement. In this regard, it was necessary to modify the nonbonded repulsion term between the Ser 195 Oγ and the scissile C to allow the Oγ to approach the carbon atom closer than the van der Waals distance. This was done by reducing the van der Waals radius of the Oγ by 35%. Additionally, the restraint for planarity of the 15(I) peptide group was removed to allow for some distortion at the scissile bond. These alterations led to a refined distance of 2.70 Å between the Ser 195 Oγ and the carbonyl carbon of 15(I) in T-APPI. Further reduction of the van der Waals radius of 195 Oγ did not lead to an additional decrease in this distance. In the C-APPI and C-BPTI complexes, the refined distances were 2.68 Å and 2.75 Å, respectively. Distances in this range have been reported in a number of similar inhibitor–protease complexes (reviewed in Bode & Huber, 1992).

Although the shorter Oγ-C distance suggests a small degree of covalent character of the bond, no apparent distortion in the planarity of the peptide bond toward a pyramidal configuration was observed (Fig. 10). This finding differs from that found in the 2.1-Å rat trypsin–APPI structure, where a distortion from planarity was reported (Perona et al., 1993a). It must be said, however, that at the resolution of all these structures, small distortions of the type suggested are beyond the limits of confident interpretation. Planar geometry at the peptide bond has been reported in a series of third domain ovomucoid inhibitor–*Streptomyces griseus* protease complexes (Read & James, 1988), whereas a very slight distortion was reported for the eglin–subtilisin complex (Bode et al., 1986).

Based on the comparison between the trypsin and chymotrypsin complexes, the stereochemistry at the site of nucleophilic attack is essentially the same for a "good" P1 side chain as for a "poor" one. However, a few caveats apply. Small, but significant, differences could be masked because the structures were done at only about 2 Å resolution; the most accurate method for determining such small differences is through difference map analysis in isomorphous crystal systems, but this could not be done in this case because each complex crystallized in a different space group. Nevertheless, it can be said that if differences do occur, they have magnitudes less than a few tenths of an Å. Additionally, analysis of the complexes shows that the presentation of the inhibitor's binding loop to the protease's active site and binding pocket involves a more restricted conformation than would exist for a free peptide substrate. For instance, the disulfide bridge at position 14, the residue preceding P1, influences the flexibility at the scissile peptide. In fact, the reason these proteins act as inhibitors rather than substrates may be due to restrictions of conformation and flexibility, properties that would not impede hydrolysis of a free substrate. Nevertheless, based on the high degree of complementarity observed for the Lys and Arg side chains in the chymotrypsin complexes, it is not clear why substrates with these P1 residues are hydrolyzed 10^4 times more slowly than "optimum" side-chain types (Hedstrom et al., 1994). It is an example of the different stereochemical criteria that are needed for simple binding versus those needed to perform efficient chemistry.

Materials and methods

Protein purification

A plasmid containing a synthetic APPI gene encoding the 58-residue sequence of APP [residues 287–344 of APP-751 (Kitaguchi et al., 1988; Ponte et al., 1988; Tanzi et al., 1988)] that shares high homology with BPTI (Castro et al., 1990) was used to transform *Escherichia coli* 27C7, a nonsuppressor derivative of *E. coli* W3110, for expression of the free, soluble inhibitor. A starter culture was grown overnight in LB broth at 37 °C and 5 mL were used to inoculate 250 mL of low-phosphate minimal medium containing

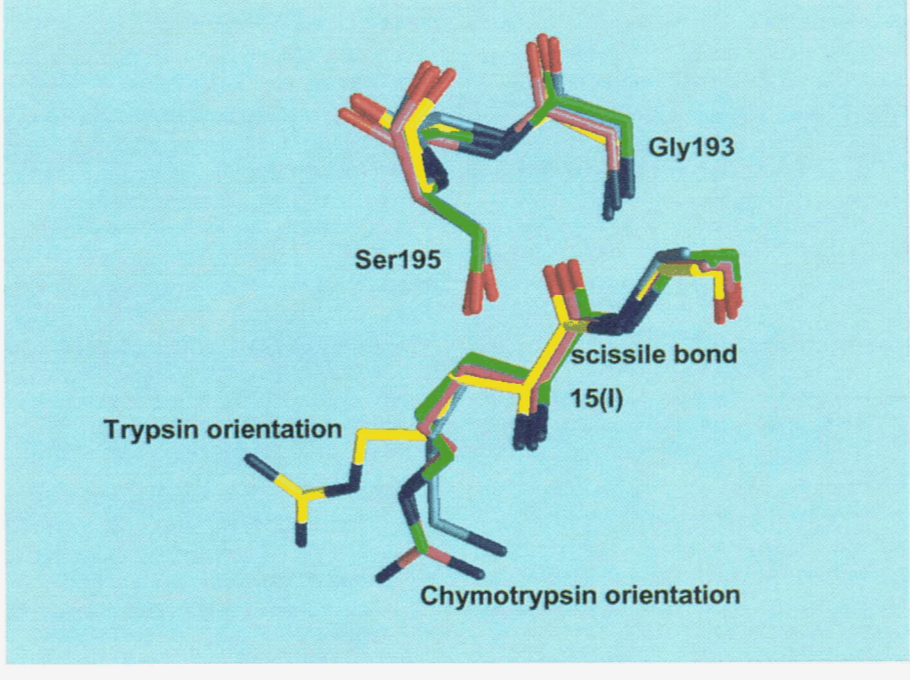

Fig. 10. Comparison of the stereochemistry of T-APPI (yellow), C-APPI (magenta), C-BPTI (green), and a mutant of C-BPTI (grey) around the scissile peptide bond.

50 μg/mL ampicillin and grown for 20 h at 37 °C. Inhibitor was secreted into the periplasm by virtue of the stII signal sequence and eventually leaked into the media (Chang et al., 1987). Cells and debris were removed by centrifugation (10,000 × *g*, 30 min). To extract the inhibitor, the supernatant was passed over an affinity column with trypsin covalently bound to Affi-Gel 10 (Bio-Rad). The inhibitor was eluted from the washed affinity column (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 20 mM $CaCl_2$) with 10 mM HCL, 500 mM KCl. The eluted inhibitor was purified by preparative reversed-phase HPLC on a Vydac C18 column (10 × 250 mm) using a 30-min linear gradient of 10–40% (v/v) acetonitrile/water/0.1% (v/v) trifluoroacetic acid (TFA) with a 2.5 mL/min flow rate. The elution profiles were monitored at both 214 nm and 280 nm. The desired protein fraction (between 28 and 32% acetonitrile) was collected and lyophilized. The proteins were verified for the proper mass using a Sciex API 3 mass spectrometer equipped with an articulated electrospray source for mass analysis. Analytical reversed-phase HPLC indicated that the affinity-purified proteins were >90% pure and the HPLC-purified samples were >99% pure. From a liter shake flask, approximately 1 mg of APPI could be purified. Fermentation in 10-liter cultures increased the yield to approximately 5 mg per liter.

Bovine trypsin (TPCK treated) was obtained from Sigma. β-Trypsin was purified away from other forms using a modification of published procedures (Schroeder & Shaw, 1968; Fehlhammer & Bode, 1975). Protein was dissolved in 10 mM Tris-HCl, pH 7.5, 20 mM $CaCl_2$, and subjected to isocratic separation at room temperature on a Mono-S ion exchange column (Pharmacia FPLC system) equilibrated with the same buffer. Excess APPI was added immediately to fractions containing β-trypsin so that complex formation would protect β-trypsin from proteolysis. Pooled fractions were concentrated using a centriprep-10 (Amicon). Buffer exchange, as well as removal of uncomplexed APPI (M_r = 6.5 kDa), was performed by dialysis into 20 mM Hepes, pH 7.2, using a dialysis membrane with a 10-kDa molecular weight cutoff.

Bovine chymotrypsin was obtained from Worthington (CDS) and BPTI from Sigma (T-8642). Both proteins were used without further purification. To form the 1:1 complex between chymotrypsin and inhibitor, 100 μL of 2.0 mM chymotrypsin in 1 mM HCl, 20 mM $CaCl_2$ were added to 120 μL of 2.0 mM APPI or BPTI in 20 mM Hepes-NaOH, pH 7.2, 10 mM $CaCl_2$, and allowed to incubate for 12 h at 4 °C. The concentration of the protein stock solutions was adjusted using the A_{280nm} reading of a 1:1,000 diluted solution [$E_{1\%280nm}$ = 20.4 (chymotrypsin), 14.3 (trypsin), 17.2 (APPI and BPTI)].

Crystallization

For crystallization of the trypsin–APPI complex, the solution was set up at 40 mg/mL in hanging drops subjected to vapor diffusion (McPherson, 1990) at 4 °C against a reservoir containing 25% (w/v) PEG 3400, 200 mM ammonium acetate, and 100 mM sodium citrate adjusted to pH 6.5. Small crystals obtained from the hanging drops were seeded into sitting drops with a protein concentration of 26 mg/mL and equilibrated against the same reservoir solution containing 23% (w/v) PEG 3400. Crystals suitable for data collection (0.2 × 0.2 × 0.4 mm) were grown from seeds in approximately one month.

Crystals of the 1:1 complex between chymotrypsin and APPI were grown at 4 °C by the hanging drop vapor diffusion method equilibrating against a reservoir solution containing 10–12.5% (w/v) PEG-6000 (Fluka), 100 mM sodium citrate, pH 4.0, and 5% (v/v) dimethylacetamide. The crystallization drop was a mixture of 10 μL complex solution and 5 μL reservoir solution. Suitable crystals grew within a week, reaching their final size of 1.0 × 0.4 × 0.2 mm in about one month.

Chymotrypsin–BPTI crystals were grown at room temperature in sitting drops by equilibrating a 35-μL drop containing 20 mg/mL of the complex, 0.5% (v/v) MPD, 8% (v/v) saturated ammonium sulfate, 80 mM Tris-HCl, pH 7.5, against a reservoir containing 50% (v/v) saturated ammonium sulfate, 80 mM Tris-HCl, pH 7.5. The final size of the crystals was 0.5 × 0.3 × 0.3 mm.

X-ray data collection

The data sets for the trypsin–APPI and the chymotrypsin–BPTI complexes were collected at room temperature on an Enraf-Nonius "FAST" area detector mounted on a Rigaku RU200 rotating Cu-anode generator operated at 45 kV, 110 mA, with a 3-mm focal spot tube and a graphite monochromator. Data frames covering a 0.1° oscillation range were exposed for 120 s. The crystals of the chymotrypsin–APPI complex were radiation sensitive. The resolution improved from 3.1 Å to 2.0 Å by flash-freezing the crystals, which also eliminated the radiation damage. A cryogenic data set was collected on beamline 7.1 at SSRL using a Mar image plate system (300-mm plate diameter) and reduced with the XDS program package (Kabsch, 1988a, 1993). The data sets collected with the "FAST" area detector were processed with the MADNES (Messerschmidt & Pflugrath, 1987) and PROCOR (Kabsch, 1988b) software packages. The statistics for the data collection are summarized in Table 1.

Structure determination and refinement

Most of the calculations were performed with X-PLOR (Brünger et al., 1987, 1990; Brünger, 1990a; Engh & Huber, 1991) and the CCP4 program suite (CCP4, 1994). For all complexes, the starting phases for the refinement were obtained by molecular replacement using the programs X-PLOR (Brünger, 1990b) and AMoRe (Navaza, 1992, 1993). The starting search model for trypsin–APPI was derived from the coordinates of the trypsin–BPTI complex (PDB entry 2PTC; Bernstein et al., 1977; Marquart et al., 1983) and consisted of all protein atoms including the refined temperature factors. The rotation-search in combination with PC-refinement and the subsequent translation-search gave one unambiguous solution. The starting search model for the chymotrypsin–BPTI complex was constructed from chymotrypsin (PDB entry 5CHA; Blevins & Tulinsky, 1985) and BPTI of trypsin–BPTI (PDB entry 2PTC; Marquart et al., 1983) by matching the core backbone atoms of chymotrypsin with the equivalent atoms of trypsin in the trypsin–BPTI structure. All protein atoms of chymotrypsin and BPTI were used, including the refined temperature factors. Rotation searches and PC-refinement using X-PLOR produced two strong solutions that were related by a twofold symmetry axis parallel to the a^* axis, consistent with a self-rotation calculation. The two rotation solutions were used in translation searches giving strong solutions only in space group $P6_1$ and not in $P6_5$. The two complexes are related by a pseudo twofold axis between the edges of BPTI molecules that form a spiral along the sixfold screw axis with chymotrypsin molecules bound on alternate edges of the strand of BPTI molecules. A large solvent channel runs along the sixfold screw axis, which contributes to the 74% solvent content of the crystal.

The molecular replacement for the chymotrypsin:APPI complex was performed using only chymotrypsin (PDB entry, 5CHA) as the search model. This search gave two clear positive peaks. The translation function resulted for each rotation solution in one significant translation peak.

The structures were refined in an iterative fashion, using cycles of X-PLOR and PROLSQ (Hendrickson, 1985) alternated with model building into SIGMAA- (Read, 1986) weighted $2F_{obs} - F_{calc}$ and $F_{obs} - F_{calc}$ electron density maps using the molecular graphics programs FRODO (Jones, 1978) and O (Jones et al., 1991). Water molecules were added to the models based on peaks with proper H-bonding geometry and 3σ density in the difference map. The program PROCHECK (Laskowski et al., 1993) was used to evaluate the stereochemical and geometric quality, as well as to highlight outliers. Chymotrypsin–APPI and chymotrypsin–BPTI were refined with the Engh and Huber (1991) protein parameters. The chymotrypsin–BPTI crystals diffracted anisotropically and a final anisotropic overall *B*-factor refinement in X-PLOR reduced the *R*-factor by 1.4%. The six anisotropic *B*-factor parameters were refined to the following values: $B11 = -2.057$, $B22 = -2.057$, $B33 = 4.1139$, $B12 = -0.0831$, $B13 = 0.0$, and $B23 = 0.0$. The statistics of the final models are listed in Table 1. To compare the analyzed structures, the unique best set of overlapping Cα-atoms used for pairwise superpositioning of the complexes was determined with the progam NEWDOME (Perry et al., 1994) with 0.5 Å for maximum core movement and a 10-Å radius of contact. Furthermore, small alterations in interresidue distances in one molecule compared to another were evaluated by the calculation of a difference distance matrix. For the comparison of the relative orientation between the inhibitor and the protease, the angle between two vectors was calculated. The vertex point of the two vectors was set on the Cα-atom of the P1 residue (inhibitor residue 15); the end point of the vectors was the midpoint of the highly conserved disulfide bridges between the inhibitor residues Cys 5 and Cys 55 on one side and the proteinase residues Cys 136 and Cys 201 on the other. These disulfide bridges are part of the very well-conserved core of the three-dimensional fold and are distant from the active site.

Phage display mutagenesis

The plasmid pA4G32 was constructed by subcloning the synthetic gene encoding the APPI sequence (Castro et al., 1990) in place of the human growth hormone sequence in the phGHam-g3 plasmid developed for the monovalent display (Bass et al., 1990; Lowman et al., 1991). The plasmid contains the alkaline phosphatase promoter, stII secretion signal, the APPI gene, and codons 249–406 of the M13 gene III protein (Fig. 7). The plasmid also carries the colE1 and phage f1 origins of replication as well as ampicillin resistance. The APPI gene covers the 58 residues of the Alzheimer's amyloid β-protein precursor (APP) (residues 287–344 of APP-751) (Ponte et al., 1988), which share homology with the basic pancreatic trypsin inhibitor (BPTI). The 58 codons of the APPI gene are followed by an amber codon (TAG), which allows APPI-gene III fusion phage to be produced in the *supE* XL1-Blue strain of *E. coli* (Bullock et al., 1987). Using site-directed mutagenesis (Kunkel et al., 1991), one Aat II site was introduced into the APPI gene and five duplicate sites outside the gene were eliminated (*Pst* I, *Aat* II, *Sca* I, and two *Bam*H I), so that the final construct contains nine unique restriction sites distributed along the APPI gene.

Library construction

Codons 11, 13, 15, 17, and 19 of APPI (residue numbering corresponds to that of BPTI) were mutated randomly using a synthetic oligonucleotide cassette. These positions correspond to the P_5, P_3, P_1, P_2', and P_4' residues, respectively, for serine protease substrates (Schechter & Berger, 1967). As defined, cleavage occurs between P1 and P1′ positions with numbered positions in the N-terminal direction and prime numbered positions in the C-terminal direction. The five codons were mutated simultaneously with the nucleotide sequence NNS (N = A/T/C/G; S = G/C) to generate all possible amino acid combinations with 32 codons. The five NNS triplets encode 3.4×10^7 possible nucleotide sequences and 3.2×10^6 possible amino acid sequences. To eliminate any background of wild-type APPI-phage, a synthetic oligonucleotide cassette coding for two stop codons (TAA) and a frame shift was introduced between the *Apa* I and *Xho* I sites at APPI codons 12 and 19. This construct was then cut at the *Pst* I and *Aat* II sites at codons 5 and 24 to insert an oligonucleotide cassette incorporating the random codons to create the starting library. The two complementary oligonucleotides, 5′-pGC GAA CAG GCG GAA **NNS** GGG **NNS** TGC **NNS** GCG **NNS** ATC **NNS** AGA TGG TAC TTC GAC GT-3′ and 5′-C GAA GTA CCA TCT **SNN** GAT **SNN** CGC **SNN** GCA **SNN** CCC **SNN** TTC CGC CTG TTC GCT GCA-3′, were annealed and the cassette was ligated into the cut vector. The ligation products were ethanol precipitated, resuspended in water, and electroporated (Dower et al., 1988) into JM101 cells to yield 9×10^6 transformants.

Binding selections

To produce phagemid particles displaying the inhibitor library, male JM101 or XL-1 Blue cells containing the plasmid library were infected with M13K07 helper phage (Vierra & Messing, 1987) at a multiplicity of infection of approximately 100 helper phage per cell. Phage particles were isolated and propagated as described (Bass et al., 1990). The random inhibitor libraries were sorted for 6–12 cycles of enrichment for binding to the target protease immobilized on Affigel-10 beads (Biorad). For the immobilization, each protease (3 mg) was reacted with 0.5 mL of bead slurry at 5 °C for 4 h in 0.1 M MOPS, pH 8.0. Bovine trypsin (Worthington) and bovine chymotrypsin (Worthington) were used as supplied.

The phage-binding reactions were performed in 200 μL containing 500 mM KCl, 0.5% BSA, 0.05% Tween-20, 20 mM Tris, pH 7.5. The solution contained approximately 1×10^{11} phagemid particles from the starting APPI–phage display library that carry ampicillin resistance. Also included were $\sim 1 \times 10^{12}$ nondisplaying phagemid particles that carry chloramphenicol resistance. The nondisplaying phage are included to reduce nonspecific binding and to monitor the enrichment of displaying relative to nondisplaying phage. For libraries that incorporate negative selection, proteases were added in solution to the binding reactions to a final concentration of 1 μM and allowed to incubate with the phage for 5 min at 25 °C before the addition of beads. Ten milliliters of a bead slurry containing the positively selected protease target was added to the binding reaction and mixed by rotation at 25 °C for 2 h.

The beads were washed six times with 1-mL volumes of the reaction buffer to remove unbound phage. Bound phage were eluted by incubation with 0.5 μL of 500 mM KCl, 10 mM HCl for 10 min. The beads were removed by centrifugation and the supernatant was neutralized. Half of the supernatant volume was com-

bined with 1 mL of male XL-1 Blue cells, incubated at 37 °C for 30 min to allow infection, and grown overnight in 25 mL of 2YT with M13K07 and ampicillin (50 mg/mL) to propagate the selected phage. The supernatant was also titered to determine the number of ampicillin- and chloramphenicol-resistant colony-forming units/μL. The 25-μL phage cultures were PEG precipitated, resuspended in 0.5 mL of STE (10 mM Tris, pH 7.6, 1 mM EDTA, 100 mM NaCl), and used in subsequent rounds of binding selections.

Starting from the random library, binding selections were performed to trypsin and chymotrypsin under four different conditions. The positively selected protease (T^+ or C^+) is bound to the beads used in the binding selection. Negative selection (T^- or C^-) was implemented by pre-incubating the library with the negative-selected protease free in solution and then selecting for inhibitors that would bind to the target protease bound to beads.

Following 6–12 rounds of binding selection, individual clones from the selected libraries were isolated and sequenced (Sanger et al., 1977). Approximately 15–20 colonies were picked at random from each selected library and ssDNA was prepared for sequencing. To determine the sequence distribution in the starting library, 27 independent clones were picked and sequenced. The calculated frequencies of amino acids in the starting library were used to correct the sequence distributions of the selected libraries.

Protein expression

The spectrum of inhibitory activity against the three proteases was determined for the populations of selected sequences expressed as a pool, as well as individual clones that were representative of the consensus sequence of each library. ssDNA prepared from the selected libraries or clones was transformed into *E. coli* strain 27C7, a nonsuppressor derivative of W3110 *ton*A (ATCC 27325), for expression (Chang et al., 1987). Inhibitor protein was secreted free in solution as directed by the stII signal sequence and transcription under the control of the alkaline phosphatase promoter. An overnight culture was inoculated (1% v:v) into 20 μL of AP5 minimal media containing carbenicillin (50 mg/mL) and grown for 24 h at 37 °C (Chang et al., 1987). Cells were removed by centrifugation and the supernatants were dialyzed into distilled water, lyophilized, and brought up in 400 μL of 0.1 M Tris, pH 7.5. The concentrated supernatants were assayed for inhibitory activity toward trypsin, chymotrypsin, and crude protease.

Inhibition assays were performed in 96-well microtiter plates at room temperature in a buffer of 0.1 M Tris, pH 7.5, 0.5 M NaCl, 20 mM $CaCl_2$, and 0.005% Triton X-100. A stock solution of trypsin was titrated with the active site inhibitor MUGB according to the method of Jameson et al. (1973). This stock was used to calibrate a solution of α1-antitrypsin inhibitor, which was used in turn to quantitate the concentrations of trypsin, chymotrypsin, and crude protease in the binding assays.

Inhibition assays

The assays of protease inhibitory activity employed samples with a fixed concentration of protease and varying concentrations of inhibitor (Green & Work, 1953; Empie & Laskowski, 1982; Kitaguchi et al., 1990). One-hundred milliliters of concentrated inhibitor supernatant (inhibitor concentration ~1–2 mm) from one of the expressed libraries or individual clones was added to a microtiter well containing 100 mL of buffer. Serial twofold dilutions were made into adjacent wells. Forty milliliters from each well in this plate was transferred to a second plate with wells containing 210 μL of buffer containing a standardized concentration of trypsin or chymotrypsin in the range of 100 nM. Each column of wells in this plate had a fixed concentration of protease and a range of inhibitor concentrations on either side of being equimolar with the protease. This high-concentration plate was used to measure inhibitor concentration. Medium- and low-concentration plates were made by diluting the high-concentration plate 10- and 100-fold, respectively, with buffer. The medium- and low-concentration plates were used to measure K_i as described previously (Dennis & Lazarus, 1994a; Seymour et al., 1994).

Following 1–3 h of incubation to allow for equilibration, enzyme activities were measured as follows. The high-concentration plates were assayed by adding 25 μL of a 2.5 mg/mL solution of a pNA substrate dissolved in DMSO and monitoring the absorbance increase at 405 nm in a SLT EAR 340 AT plate reader. pNA substrates for trypsin, chymotrypsin, and crude protease were BAPA, Ala-Ala-Pro-Leu-pNA, and Ala-Ala-Pro-Phe-pNA, respectively. Enzyme activities in the medium- and low-concentration plates were assayed by adding 25 μL of a 0.5 mg/mL solution of a MCA substrate dissolved in DMSO and monitoring the fluorescence increase (excitation 355 nm, emission 460 nm) using a Labsystems Fluoroskan II plate reader. Data were collected using Delta Soft II SLT or FL v3.31 software (BioMetallics). MCA substrates for trypsin, chymotrypsin, and crude protease were Boc-Ile-Glu-Gly-Arg-MCA, N-Succ-Ala-Ala-Pro-Phe-MCA, and Pro-Phe-Arg-MCA, respectively. Nonlinear regression analysis of the data was performed as described previously (Seymour et al., 1994).

Acknowledgments

We thank the Organic Chemistry Department for oligonucleotide synthesis, Han Chen for fermentations, and John Stults for mass spectrometry. We thank Steve Bass and Henry Lowman for help with the phage display methods, and Robert Lazarus and Mark Dennis for protocols and advice on protease inhibitor assays. Mike Randal, Will Somers, Yves Muller, and Bart de Vos helped with data collection at SSRL. We are grateful to David Wood for preparation of figures. We acknowledge the support of SSRL for synchrotron data collection. Figures were made with GRASP (Nicholls et al., 1993). The coordinates of the complexes are deposited in the Protein Data Bank: T-APPI (1taw), C-BPTI (1cbw), C-APPI (1ca0).

References

Barrett AJ, Rawlings ND. 1995. Families and clans of serine peptidases. [Review]. *Arch Biochem Biophys 318*:247–250.

Bass S, Greene R, Wells JA. 1990. Hormone phage: An enrichment method for variant proteins with altered binding properties. *Proteins Struct Funct Genet 8*:309–314.

Bauer CA, Thompson RC, Blout ER. 1976. The active centers of *Streptomyces griseus* protease 3, α-chymotrypsin, and elastase: Enzyme–substrate interactions close to the scissile bond. *Biochemistry 15*:1296–1299.

Baumann WK, Bizzozero SA, Dutler H. 1970. Specificity of α-chymotrypsin. Dipeptide substrates. *FEBS Lett 8*:257.

Berezin IV, Martinek K. 1970. Specificity of α-chymotrypsin. *FEBS Lett 8*:261.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Blevins RA, Tulinsky A. 1985. The refinement and the structure of the dimer of α-chymotrypsin at 1.67 Å resolution. *J Biol Chem 260*:4264–4275.

Blow DM, Wright CS, Kukla D, Rühlmann AR, Steigemann W, Huber R. 1972. A model for the association of bovine pancreatic trypsin inhibitor with chymotrypsin and trypsin. *J Mol Biol 69*:137–144.

Bode W, Huber R. 1992. Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem 204*:433–451.

Bode W, Papamokos E, Musil D, Seemueller U, Fritz H. 1986. Refined 1.2 Å crystal structure of the complex formed between subtilisin Carlsberg and the inhibitor eglin c. Molecular structure of eglin and its detailed interaction with subtilisin. *EMBO J 5*:813–818.

Brünger AT. 1990a. *X-PLOR: A system for crystallography and NMR.* New Haven, Connecticut: Yale University Press.

Brünger AT. 1990b. Extension of molecular replacement: A new search strategy based on Patterson correlation refinement. *Acta Crystallogr A 46*:46–57.

Brünger AT, Krukowski A, Erickson JW. 1990. Slow cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr A 46*:583–593.

Brünger AT, Kuriyan J, Karplus M. 1987. Crystallographic R-factor refinement by molecular dynamics. *Science 235*:458–460.

Bullock WO, Fernandez JM, Short JM. 1987. XL1-Blue: A high efficiency plasmid transforming recA *Escherichia coli* strain with beta-galactosidase selection. *Bio-Techniques 5*:376–379.

Castro M, Marks CB, Nilsson B, Anderson S. 1990. Does the Kunitz domain from the Alzheimer's amyloid beta protein precursor inhibit a kallikrein responsible for post-translational processing of nerve growth factor precursor? *FEBS Lett 267*:207–212.

Castro MJM, Anderson S. 1996. Alanine point-mutations in the reactive region of bovine pancreatic trypsin inhibitor (BPTI): Effects on the kinetics and thermodynamics of binding to β-trypsin and α-chymotrypsin. *Biochemistry 35*:11435–11446.

CCP4 (1994). The CPP4 suite: Programs for protein crystallography. *Acta Cryst D 50*:760–763.

Chang CN, Rey M, Bochner B, Heyneker H, Gray G. 1987. High-level secretion of human growth hormone by *Escherichia coli*. *Gene 55*:189–196.

Creighton TE, Charles IG. 1987. Biosynthesis, processing, and evolution of bovine pancreatic trypsin inhibitor. *Cold Spring Harbor Symp Quant Biol 52*:511–519.

Creighton TE, Darby NJ. 1989. Functional evolutionary divergence of proteolytic enzymes and their inhibitors. *Trends Biochem Sci 14*:319–324.

Dennis MS, Herzka A, Lazarus RA. 1995. Potent and selective Kunitz domain inhibitors of plasma kallikrein designed by phage display. *J Biol Chem 270*:25411–25417.

Dennis MS, Lazarus RA. 1994a. Kunitz domain inhibitors of tissue factor-factor VIIa: Potent inhibitors selected from libraries by phage display. *J Biol Chem 35*:22129–22136.

Dennis MS, Lazarus RA. 1994b. Kunitz domain inhibitors of tissue factor-factor VIIa: Potent and specific inhibitors by competitive phage selection. *J Biol Chem 35*:22137–22144.

Dower WJ, Miller JF, Ragsdale CW. 1988. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res 16*:6127.

Empie MW, Laskowski M. 1982. Thermodynamics and kinetics of single residue replacements in avian ovomucoid third domains: Effect on inhibitor interactions with serine proteinases. *Biochemistry 21*:2274–2284.

Engh RA, Huber R. 1991. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr A 47*:392–400.

Evnin LB, Vasquez JR, Craik CS. 1990. Substrate specificity of trypsin investigated by using a genetic selection. *Proc Natl Acad Sci USA 87*:6659–6663.

Fehlhammer H, Bode W. 1975. The refined crystal structure of bovine beta-trypsin at 1.8 Å resolution. I. Crystallization, data collection and application of Patterson search technique. *J Mol Biol 98*:683–692.

Graf LG, Hegyi I, Liko J, Hepp K, Medzihradszky C, Craik C, Rutter WJ. 1988. Structural and functional integrity of specificity and catalytic sites of trypsin. *Int J Peptide Protein Res 32*:512–518.

Green NM, Work E. 1953. Pancreatic trypsin inhibitor: Reaction with trypsin. *Biochem J 54*:347–352.

Hedstrom LL, Perona JJ, Rutter WJ. 1994. Converting trypsin to chymotrypsin: Residue 172 is a substrate specificity determinant. *Biochemistry 33*:8757–8763.

Hedstrom LL, Szilagyi, Rutter WJ. 1992. Converting trypsin to chymotrypsin: The role of surface loops. *Science 255*:1249–1253.

Hendrickson WA. 1985. Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol 115*:252–270.

Huber R, Kukla D, Bode W, Schwager P, Bartels K, Deisenhofer J, Steigemann W. 1974. Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. *J Mol Biol 89*:73–101.

Hynes TR, Randal M, Kennedy LA, Eigenbrot C, Kossiakoff AA. 1990. X-ray crystal structure of the protease inhibitor domain of Alzheimer's amyloid β-protein precursor. *Biochemistry 29*:10018–10022.

Jameson GW, Roberts DV, Adams RW, Kyle WSA, Elmore DT. 1973. Determination of the operational molarity of solutions of bovine a-chymotrypsin, trypsin, thrombin and factor Xa by spectrofluorimetric titration. *Biochem J 131*:107–117.

Jones AT. 1978. A graphics model building and refinement system for macromolecules. *J Appl Crystallogr 11*:268–272.

Jones AT, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A 47*:110–119.

Kabsch W. 1988a. Automatic indexing of rotation diffraction patterns. *J Appl Crystallogr 21*:67–71.

Kabsch W. 1988b. Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J Appl Crystallogr 21*:916–934.

Kabsch W. 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr 26*:795–800.

Keil B. 1992. In: *Specificity of proteolysis.* Heidelberg, Berlin: Springer-Verlag. pp 66–110.

Kitaguchi N, Takahashi Y, Oishi K, Shiojiri S, Tokushhima Y, Utsunomiya T, Ito H. 1990. Enzyme specificity of proteinase inhibitor region in amyloid precursor protein of Alzheimer's disease: Different properties compared with protease nexin I. *Biochim Biophys Acta 1038*:105–113.

Kitaguchi N, Takahashi Y, Tokushhima Y, Shiojiri S, Ito H. 1988. Novel precursor of Alzheimer's disease amyloid protein shows protease inhibitory activity. *Nature 331*:530–532.

Kossiakoff AA. 1987. Catalytic properties of trypsin. In: Jurnak, FA, McPherson A, eds. *Biological macromolecules and assemblies, vol 3: Active sites of enzymes.* New York: John Wiley & Sons. pp 369–412.

Kunkel TA, Bebenek K, McClary J. 1991. Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol 204*:125–139.

Laskowski M Jr, Kato I. 1980. Protein inhibitors of proteinases. *Annu Rev Biochem 49*:593–626.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr 26*:283–291.

Lowman HB, Bass SH, Simpson N, Wells JA. 1991. Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry 30*:10832–10838.

Lu W, Apostol I, Qasim MA, Warne N, Wynn R, Zhang WL, Anderson S, Chiang YW, Ogin E, Rothberg I, Ryan K, Laskowski M. 1997. Binding of amino acid side chains to S1 cavities of serine proteinases. *J Mol Biol 266*:441–461.

Markland W, Ley AC, Lee SW, Ladner RC. 1996. Iterative optimization of high-affinity protease inhibitors using phage display. 1. Plasmin. *Biochemistry 35*:8045–8057.

Marquart M, Walter J, Deisenhofer J, Bode W, Huber R. 1983. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr B 39*:480.

McPherson A. 1990. Current approaches to macromolecular crystallization. *Eur J Biochem 189*:1–23.

Messerschmidt A, Pflugrath JW. 1987. Crystal orientation and X-ray pattern prediction routines for area-detector diffractometer systems in macromolecular crystallography. *J Appl Crystallogr 20*:306–315.

Navaza J. 1992. AMoRe. A new package for molecular replacement: Application to some difficult cases: Anonymous molecular replacement. *Proc Daresbury Study Weekend.* Daresbury, UK: Daresbury Laboratory.

Navaza J. 1993. On the computation of the fast rotation function. *Acta Crystallogr D 49*:588–591.

Neurath H. 1984. Evolution of proteolytic enzymes. [Review]. *Science 224*:350–357.

Nicholls AR, Bharadwaj, Honig B. 1993. GRASP. Graphical representation and analysis of surface properties. *Biophys J 64*:A166.

Perona JJ, Craik CS. 1995. Structural basis of substrate specificity in the serine proteases. *Protein Sci 4*:337–360.

Perona JJ, Tsu CA, Craik CS, Fletterick RJ. 1993a. Crystal structures of rat anionic trypsin complexed with the protein inhibitors APPI and BPTI. *J Mol Biol 230*:919–933.

Perona JJ, Luke BE, Craik CS. 1993b. A generic selection elucidates structural determinants of arginine verses lysine specificity in trypsin. *Gene 137*:121–126.

Perry KM, Fauman EB, Finer-Moore JS, Montfort WR, Maley GF, Maley F, Stroud RM. 1994. Plastic adaptation toward mutations in proteins: Structural comparisons of thymidylate synthases. *Proteins Struct Funct Genet 8*:315–333.

Ponte P, Gonzalez-DeWhitt P, Schilling J, Miller J, Hsu D, Greenberg B, Davis K, Wallace W, Lieberburg I, Fuller F, Cordell B. 1988. A new A4 amyloid mRNA contains a domain homologous to serine proteinase inhibitors. *Nature 331*:525–527.

Read RJ. 1986. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr A 42*:140–149.

Read RJ, James MN. 1986. Introduction to protein inhibitors: X-ray crystallography. In: Barrett AJ, Salvesen G, eds. *Protease inhibitors.* Amsterdam: Elsevier. pp 301–336.

Read RJ, James MN. 1988. Refined crystal structure of *Streptomyces griseus* trypsin at 1.7 Å resolution. *J Mol Biol 200*:523–551.

Roberts BL, Markland W, Ley AC, Kent RB, White DW, Guterman SK, Ladner RC. 1992. Directed evolution of a protein: Selection of potent neutrophil

elastase inhibitors displayed on M13 fusion phage. *Proc Natl Acad Sci USA 89*:2429–2433.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA 74*:5463–5467.

Schechter I, Berger M. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun 27*:157–162.

Schroeder DD, Shaw E. 1968. Chromatography of trypsin and its derivatives: Characterization of a new active form of bovine trypsin. *J Biol Chem 243*:2943–2949.

Seymour JL, Lindquist RN, Dennis MS, Moffat B, Yansura D, Reilly D, Wessinger ME, Lazarus RA. 1994. Ecotin is a potent anticoagulant and reversible tight-binding inhibitor of factor Xa. *Biochemistry 33*:3949–3958.

Stroud RM. 1974. A family of protein-cutting proteins. *Sci Am 231*:74–88.

Tanzi RE, McClatchey AI, Lamperti ED, Villa-Komaroff L, Gusella JF, Neve RL. 1988. Protease inhibitor domain encoded by an amyloid protein precursor mRNA associated with Alzheimer's disease. *Nature 331*:528–530.

Vierra J, Messing J. 1987. Production of single-stranded plasmid DNA. *Methods Enzymol 153*:3–11.

Wagner SL, Siegel S, Vedvick TS, Raschke WC, Van Nostrand WE. 1992. High level expression, purification, and characterization of the Kunitz-type protease inhibitor domain of protease nexin-2/amyloid beta-protein precursor. *Biochem Biophys Res Commun 186*:1138–1145.

Wells JA, Lowman HB. 1992. Rapid evolution of peptide and protein binding properties in vitro. *Curr Opin Struct Biol 2*:597–604.