doi:10.1006/jmbi.2001.5305 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 315, 1209–1218

JMB

Crystal Structure Reveals Basis for the Inhibitor Resistance of Human Brain Trypsin

Gergely Katona[1,2], Gunnar I. Berglund[2], Janos Hajdu[2], László Gráf[1] and László Szilágyi[1]*

[1]*Eötvös Loránd University Department of Biochemistry Puskin u. 3., 1088 Budapest Hungary*

[2]*Uppsala University Department of Biochemistry P.O. Box 576, S-751 23 Uppsala, Sweden*

**Corresponding author*

Severe neurodegradative brain diseases, like Alzheimer, are tightly linked with proteolytic activity in the human brain. Proteinases expressed in the brain, such as human trypsin IV, are likely to be involved in the pathomechanism of these diseases. The observation of amyloid formed in the brain of transgenic mice expressing human trypsin IV supports this hypothesis. Human trypsin IV is also resistant towards all studied naturally occurring polypeptide inhibitors. It has been postulated that the substitution of Gly193 to arginine is responsible for this inhibitor resistance. Here we report the X-ray structure of human trypsin IV in complex with the inhibitor benzamidine at 1.7 Å resolution. The overall fold of human trypsin IV is similar to human trypsin I, with a root-mean square deviation of only 0.5 Å for all C^{α} positions. The crystal structure reveals the orientation of the side-chain of Arg193, which occupies an extended conformation and fills the S2′ subsite. An analysis of surface electrostatic potentials shows an unusually strong clustering of positive charges around the primary specificity pocket, to which the side-chain of Arg193 also contributes. These unique features of the crystal structure provide a structural basis for the enhanced inhibitor resistance, and enhanced substrate restriction, of human trypsin IV.

Keywords: protein crystallography; inhibitor resistance; brain trypsin; hereditary pancreatitis; Alzheimer disease

Introduction

Trypsins belong to the superfamily of serine proteinases and were first isolated from the pancreatic juice of animals, but later were identified in many different tissues. Their predominant activity is the digestion of food in the duodenum. Secondary functional roles have been identified as the activation of cell surface receptors (PAR family);[1] in processing polypeptide hormone precursors; and in the digestion of proteins of the extracellular matrix, thus helping the migration of cells.[2] The mRNA corresponding to human trypsinogen IV was the only trypsin-related mRNA found in the brain, and it showed strong tissue specificity. Its alternatively spliced isoform, named mesotrypsinogen, is expressed in low amounts in the pancreas. The two isoforms (both termed A-form) differ only in the first exon that encodes the signal peptide but otherwise their sequences (including the putative propeptide) are identical. The gene encoding these isoforms is located on chromosome 9, whereas those for human trypsin I and II are located on chromosome 7.[3] This separation of trypsinogen loci is reported only in the human genome, suggesting that human trypsin IV might be characteristic to Primates. A polymorph variant of human trypsinogen IV (termed B-form) was also identified, which lacks one glutamate residue from the N-terminal

Present address: G. Katona, Chalmers University of Technology, Department of Molecular Biotechnology, P.O. Box 462 SE-405 30 Göteborg, Sweden.

Abbreviations used: Cbz-VGR-pNA, carboxybenzoyl-valinyl-glycyl-arginyl-*p*-nitro-anilide; Cbz-K-Sbzl, Nα-carboxybenzoyl-lysyl-thiobenzyl ester; Suc-AAPR-AMC, succinyl-alanyl-alanyl-prolyl-arginyl-aminomethyl coumarin; Suc-AAPK-AMC, succinyl-alanyl-alanyl-prolyl-lysyl-aminomethyl coumarin; BPTI, bovine pancreatic trypsin inhibitor; APPI, Alzheimer precursor protein trypsin inhibitor domain; SBTI, soybean trypsin inhibitor; PEG, polyethylen glycol; bt, bovine trypsin; ht, human trypsin; rt, rat trypsin.

E-mail address of the corresponding author: szilagyi@cerberus.elte.hu

0022-2836/02/051209–10 $35.00/0

loop (residue 23 or 24)†.[4] This study is concerned with the A-form of human trypsin IV.

The predominant function of human trypsin IV in the human brain is not fully understood. A possible involvement in pathological and physiological processes is suggested by transgenic experiments, which have shown that the overexpression of human trypsinogen IV in the neurons of mice causes enhanced expression level of glial fibrillar acidic protein, and also an increase in the amount of amyloid fragments in the astrocytes.[5] Furthermore, a series of studies have shown that human trypsin IV consistently displays an unusually high resistance against all naturally occurring inhibitors to be assayed.[4,6,7] Nevertheless, studies to date have not provided a comprehensive survey of binding affinities over a number of naturally occurring inhibitors. Furthermore, although it has been suggested that this increased resistance may have a structural basis, there is no published structure of a trypsin containing the Gly193 to Arg substitution characteristic of human trypsin IV. Here, we address this shortcoming by determining the X-ray structure of human trypsin IV (in complex with benzamidine) at high resolution. We also measure the dissociation constants for several naturally occurring proteinase inhibitors, and correlate these kinetic measurements to the recovered X-ray structure through comparison with the published structures of other trypsin:inhibitor complexes. The combined approach yields a structural basis for the enhanced inhibitor resistance of human trypsin IV.

Results

X-ray structure of human trypsin IV

Crystals of human trypsin IV in complex with benzamidine were grown (see Materials and Methods) and diffracted to 1.7 Å resolution. Crystals were grown in the presence of the inhibitor benzamidine so as to avoid problems associated with autolysis. These crystals belong to space group $P4_12_12$ with one molecule per asymmetric unit. The structure was solved by molecular replacement using the X-ray structure of human trypsin I as a starting model (see Materials and Methods). X-ray diffraction data and final refinement statistics are summarized in Table 1. The crystallographic *R* factor of the refined model of benzamidine:human trypsin IV complex is 0.188, with a free *R* factor of 0.203. The model is well defined in electron density except for a few side-chains of residues located in surface loops (24, 62, 135, 165, 169, 175, 178 and 222), and 125 water molecules were included in the final structure.

Human trypsin IV displays a high structural homology with other known trypsins. Superposition of 223 C^α atoms of human trypsin IV onto human trypsin I, bovine cationic trypsin, and rat anionic trypsin yielded a root mean square deviation of C^α 0.50, 0.42 and 0.33 Å, respectively. The high level of structural similarity is largely anticipated on the basis of high sequence homology: 87.5 % to human trypsin I, 78.0 % to rat anionic trypsin and 72.6 % to bovine trypsin. Human trypsin I is an exception because even though the sequence is highly homologous to human trypsin IV (87.5 %), the structure is the least similar to human trypsin IV based on the root mean square deviation of C^α positions (0.50 Å). As shown in Figure 1, the largest differences occur near the loop regions. Deviations of more than 1.0 Å are observed around residues 23-26, 76 and 149 relative to human trypsin I; around residues 129 and 178 relative to rat trypsin; and near residues 23, 61-62, 116, 125, 129, 147-148 relative to bovine trypsin. A rationale for the structural changes observed due to the substitution of Gly193 to Arg193; near the N-terminal loop (residues 23 to 26); and those due to the loss of the disulphide bond (between residues 128 and 232), are described in detail below.

Arginine 193

A unique feature of human trypsin IV is the replacement of the otherwise conserved Gly193 by an arginine (Figure 2). This substitution is of particular significance since, in both cases, the amido nitrogen acts as a hydrogen bond donor within the oxyanion hole. A small shift, between 0.4 and 0.8 Å, is seen at the C^α position of Arg193. Nevertheless, the specific geometry of the oxyanion hole is well preserved. As such this small distortion to the oxyanion hole, a feature characteristic of the family of serine proteases, is unlikely to influence the catalytic activity of human trypsin IV. Indeed,

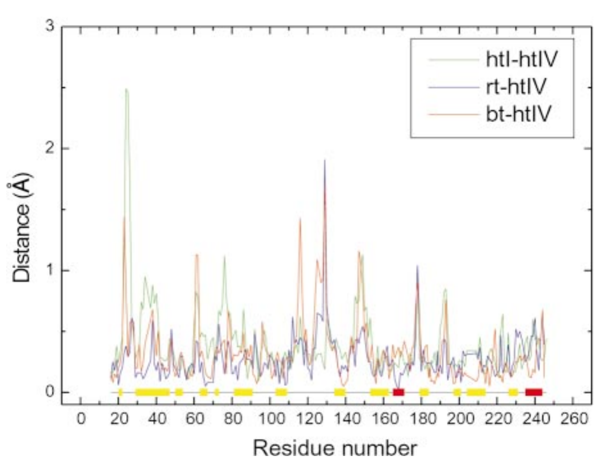

Figure 1. Distance plot. Distances are displayed between corresponding C^α atoms after superposition of human I, bovine and rat trypsin on human trypsin IV. The secondary structure is indicated on the bottom line, the beta sheets (yellow blocks) and helices (red blocks) are connected with loop regions (thin black line).

† The conventional chymotrypsin amino acid numbering is used throughout.[39]

Table 1. Data collection and refinement statistics

A. *Data collection*[a]			
Space group		$P4_12_12$	
Cell parameters		$a=b=56.68$ Å, $c=143.21$ Å	
No. of reflections		111,768	
No. of unique reflections		26,174	
Resolution range (Å)		52.7-1.7 (1.74-1.70)	
Mosaicity		0.1	
R_{sym}[b]		0.049 (0.323)	
Multiplicity		4.3	
$\langle I\rangle/\langle\sigma(I)\rangle$		18.5 (2.5)	
Completeness (%)		94.9 (92.0)	
B. *Refinement*			
R factor (%)[c]		18.8	
Free R factor (%)[29]		20.3	
No. of unique reflections		26,170	
No. of non-hydrogen atoms			
Protein		1893	
Ligand		10	
Solvent		125	
R.m.s.d. bonds (Å)/angles (deg.)		0.004/1.3	
Average B-factors ± standard deviations (Å^2)	Protein	Waters	Overall
Main-chain	17.5 ± 5.9		
Side-chain	20.7 ± 7.0		
All	19.0 ± 6.2	31.1 ± 9.5	29.8 ± 7.5

[a] Values in parentheses indicate statistics for the highest resolution shell.
[b] $R_{sym} = \Sigma|I_o - \langle I\rangle|/\Sigma I_o \times 100\%$, where I_o is the observed intensity of a reflection and $\langle I\rangle$ is the average intensity obtained from multiple observations of symmetry related reflections.
[c] R factor $= \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}| \times 100\%$.

in the kinetic studies on a synthetic substrate described below the rate of catalysis was equivalent for both human trypsin IV and bovine trypsin.

Because of the unique nature of the arginine at position 193, the conformation of its side-chain is highly influential in determining the binding affinity of substrates and inhibitors. Despite the fact that this side-chain was located on the protein's surface, its electron density was well ordered and its specific conformation could be assigned unambiguously (Figure 3). The orientation of the charged guanidium group points away from the S1 specificity pocket. The side-chain adopts an extended conformation, which lies in a cleft deli-

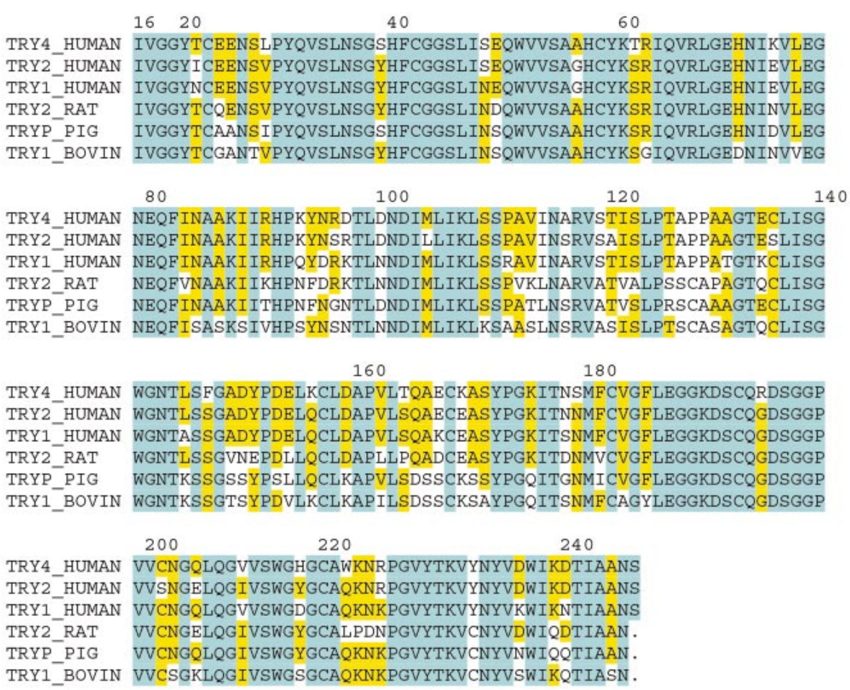

Figure 2. Sequence alignment of human trypsin IV and related trypsins. Green background shows a fully conserved amino acid position, yellow marks the majority of a specific residue at a given position.

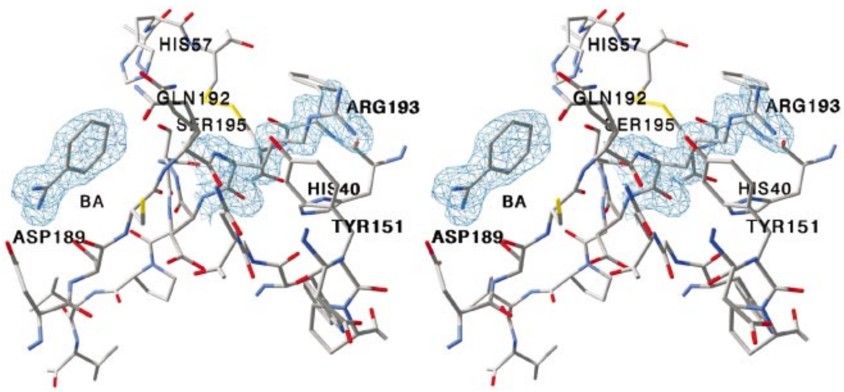

Figure 3. Stereo view of Arg193 and neighbouring residues (39-41, 141-142, 150-152). Residues between 189-200, His57 and the benzamidine (BA) molecule in the substrate binding pocket are also displayed. The SigmaA weighted $2mF_o - DF_c$ electron density map around Arg193 and benzamidine is contoured at 1σ level. Gln192 is shown in the two alternative conformations.

neated by the side-chains of Tyr151, His40, and the backbone atoms of Phe41, Trp141 and Gly142. The aromatic ring of Tyr151 interferes with the free rotation of the side-chain of Arg193, and may help to keep this residue well ordered.

As would be expected, the introduction of a new charged group greatly influences the surface electrostatic potentials. In Figure 4 the surface charge distributions for human trypsin IV, human trypsin I, rat anionic trypsin and bovine cationic trypsin are compared. All trypsins share the negatively charged primary substrate binding pocket (S1 pocket). This pocket gives trypsin its substrate specificity,[8] due to the presence of Asp189, which interacts with the positively charged side-chain of an arginine or a lysine. Nevertheless, the distribution of charges along the substrate binding cleft varies markedly. For human trypsin I this pocket is surrounded by negative charges; the charge distribution for rat anionic trypsin and bovine cationic trypsin is somewhat less polarized, whereas human trypsin IV shows two clusters of positive charges in the local neighbourhood of the S1 substrate binding pocket. This modification of surface potentials is especially apparent at the S2′ subsite: in human trypsin I the presence of an anionic modifying group (presumably a phosphate) covalently bound to Tyr151[9] contributes to the subsite's negative polarity, whereas this site is predominantly positive in human trypsin IV because of Arg193. In addition, a second reversal of polarity is also visible close to the substrate binding site: position 217 is occupied by an aspartate in human trypsin I, whereas this is a positively charged histidine in human trypsin IV. Such dramatic changes in the charge distribution near the S1 binding pocket can be expected to strongly influence inhibitor binding. This possibility was quantified through comparative inhibitor binding studies as detailed below.

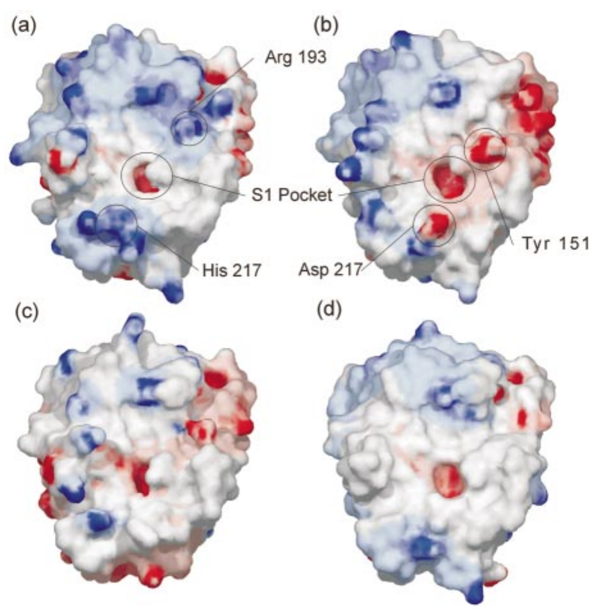

Figure 4. Electrostatic potential mapped onto the molecular surface of (a) human trypsin IV, (b) human trypsin I, (c) rat anionic trypsin and (d) bovine cationic trypsin. The contouring level of electrostatic potential is -18 kT/e (red) and 18 kT/e (blue). The orientation of the molecules is the same as in Figure 1(a).

N-terminal loop (threonine 21)

As shown in Figure 1, the largest structural differences between human trypsins I and IV are found near the N-terminal loop (positions 23-26). These conformational differences are probably caused by the substitution of Asn21 in human trypsin I by a threonine in human trypsin IV (Figure 2). This interpretation finds support from the fact that the N-terminal of both bovine trypsin and anionic rat trypsin, which also have a threonine at this position, adopt a conformation closely similar to that of human trypsin IV (Figure 5). This finding could not be anticipated from sequence comparisons alone, since human trypsins I and IV display high sequence identity in this region (Figure 2), with the only difference from residue 16 to 26 being at position 21. Additional substitutions are seen for both rat anionic trypsin (residue 23) and bovine cationic trypsin (residues 23, 24 and 26), yet both share the same conformation of the N-terminal loop as human trypsin IV.

A functional role of the N-terminal loop, and in particular residue 21, is implied by the observation that a hereditary pancreatitis is associated with the

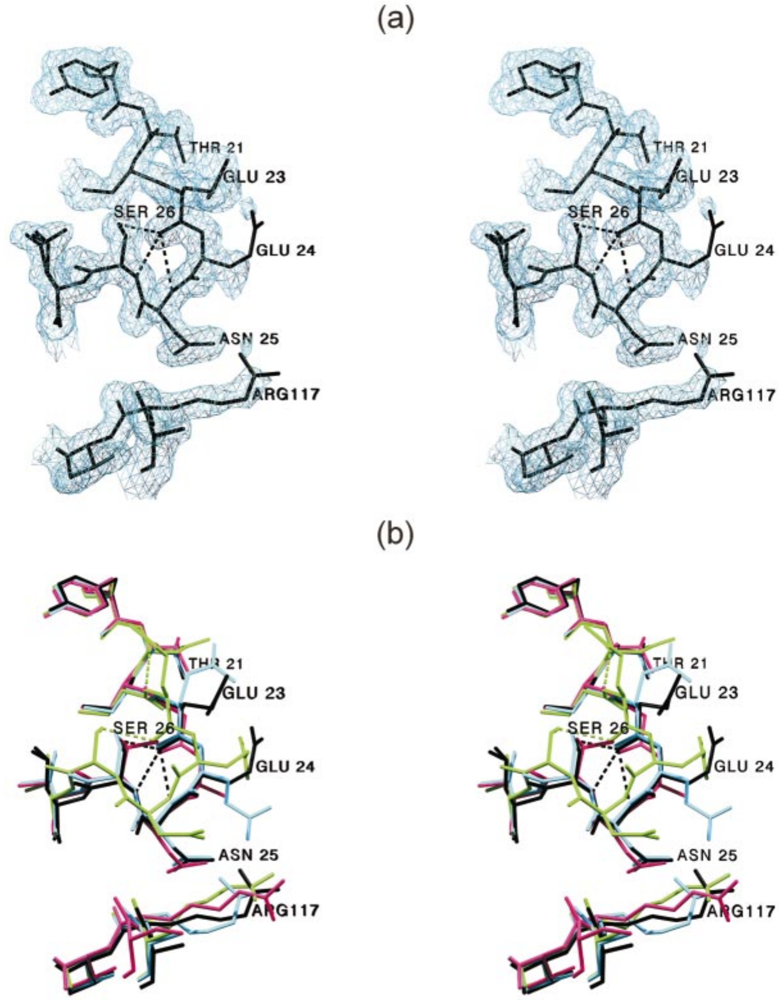

Figure 5. (a) View of the N-terminal loop of human trypsin IV. The SigmaA weighted $2mF_o - DF_c$ electron density map is contoured at 1σ level. (b) Structural comparison of the N-terminal loops in human trypsin IV (black), human trypsin I (green), rat trypsin (cyan), bovine trypsin (magenta). Two alternative conformations are shown for Leu27.

Asn21 → Ile mutation.[10] Furthermore, this site affects the rate of autoactivation of the zymogen, since the Asn21 → Thr mutation of human trypsinogen I resulted in faster autoactivation and increased trypsin stability.[11,12] Our results imply a structural basis for these observations, which could further be explored through structural studies of the Asn21 → Ile and Asn21 → Thr mutants of human trypsinogen and trypsin I.

Loss of disulfide bridge 128-232

In addition to the structural changes associated with Arg193 and Thr21, there is also a significant structural rearrangement at position 129 for the two human trypsins relative to bovine cationic trypsin. This difference may be accounted for by the lack of the 128-232 disulfiphide bridge and the presence of Pro127 and Pro128 in the human trypsins. For human trypsin I and IV the side-chain of Tyr232 (Cys232 for all other vertebrate trypsins) lies parallel to the peptide bond between residues Ala127 and Pro128. Favourable interactions between the π-electrons of the aromatic ring of Tyr232 and the peptide bond may partially compensate for the loss of the disulphide bridge, thereby maintaining the stability of the conserved fold.

Kinetic studies and modeling

The X-ray structure of human trypsin IV showed that the substitution Gly193 to an arginine had a dramatic effect on the distribution of surface charges (Figure 4). It is probable that this redistribution of surface charges would be a key factor in determining inhibitor binding strengths. This possibility was quantified by first determining the K_I values for a number of naturally occurring inhibitors, and then modeling the inhibitor bound at the active site of human trypsin IV. The correlation of these results provides a structural basis for the increased inhibitor resistance of human trypsin IV.

Catalytic activity

Spectroscopic measurements showed that, upon the addition of a synthetic substrate (Cbz-VGR-pNA) in solution, recombinant human trypsin IV was active and cleaved the peptide bond at a similar rate to bovine cationic trypsin. The k_{cat} values

Table 2. Comparison of inhibitory constants and calculated binding energies for human trypsin IV and bovine cationic trypsin

	Human trypsin IV		Bovine trypsin	
	K_I (M)	ΔG (kJ/mol)	K_I (M)	ΔG (kJ/mol)
Benzamidine	2.2×10^{-5}	27.6	2.6×10^{-5}	27.2
BPTI	1.2×10^{-6}	35.1	5.0×10^{-13} [36]	73.0
SBTI	4.2×10^{-7}	37.8	1.0×10^{-11} [37]	65.3
APPI	3.0×10^{-8}	44.6	1.1×10^{-10} [38]	59.1

for this substrate were 280 s^{-1} and 290 s^{-1} for human trypsin IV and bovine trypsin, respectively. For human trypsin IV the K_M value was three times lower than that of bovine trypsin (4.0×10^{-5} M *versus* 1.4×10^{-4} M), yielding a slightly higher catalytic efficiency (k_{cat}/K_M). Similar results were obtained with other synthetic substrates such as Cbz-K-SBzl, Suc-AAPR-AMC and Suc-AAPK-AMC. These kinetic results show that the Gly193 → Arg mutation did not affect the rate of peptide bond cleavage for these synthetic substrates, which may have been anticipated on the basis of the X-ray structure, since the specific geometry of the catalytic triad and the oxyanion hole as well as of subsites S1-S4 was well preserved.

Inhibitor binding assays and modeling

In addition to measuring the inhibitor binding strength of the synthetic inhibitor benzamidine (with which human trypsin was cocrystallised), the binding affinities of three natural inhibitors (Alzheimer precursor protein trypsin inhibitor domain, APPI; bovine pancreatic trypsin inhibitor, BPTI; and soybean trypsin inhibitor, SBTI) were also determined for human trypsin IV using progress curve analysis (see Materials and Methods). K_I values are summarized in Table 2. The rational for this choice of natural inhibitors is that structures of APPI and BPTI in complex with bovine cation trypsin[13,14] (entries 1taw and 3dtk of the Protein Data Bank) and SBTI in complex with porcine pancreatic trypsin[15] (entry 1avw) have been reported, and hence provide a basis for structural modeling. The results from the analysis of these trypsin:inhibitor structures are summarized in Table 3.

Benzamidine inhibited human trypsin IV and bovine cationic trypsin with a similar K_I, (2.2×10^{-5} and 2.6×10^{-5}, respectively). This may have been anticipated from the X-ray structure of the human trypsin IV:benzamidine complex (Figure 3), which showed that this small inhibitor sits deep in the primary substrate binding pocket. This position and orientation is very similar to that adopted by benzamidine when bound to bovine cationic trypsin[16] (entry 2ptn of the Protein Data Bank).

Of the three naturally occurring inhibitors to be studied, APPI showed the strongest affinity ($K_I = 3.0 \times 10^{-8}$ M) for human trypsin IV, and also the lowest reduction in binding strength (approximately 100-fold) relative to bovine cationic trypsin. This result is highly significant since APPI, a domain of Alzheimer precursor protein, which is expressed in the human brain, may target human brain trypsin IV.[5] A structural basis for this lowering of the binding affinity could be inferred by comparing the interface region of human trypsin IV with the published structure of the bovine cationic trypsin:APPI complex[13] (entry 1taw of the Protein Data Bank). In the bovine trypsin:APPI complex a total number of 21 residues are in contact with (i.e. lie closer than 4 Å from) the inhibitor. Of these 21 residues two are substituted (Tyr39 → Ser; Gly193 → Arg) in human trypsin IV. The substitution Tyr39 → Ser leads to the loss of a hydrogen bond between Tyr39 OH and Ser19I NH, and would be expected to lower the binding affinity. However, the Gly193 → Arg substitution, unique to human trypsin IV, appears even more significant. This is because the conformation adopted by the side-chain of Arg193 in human trypsin IV (Figure 3) would introduce a steric clash with Met17I of APPI. Clearly either the protein or peptide inhibitor must adopt a markedly different conformation in the real complex, introducing an additional energetic cost.

Table 3. Analysis of published proteinase-proteinase inhibitor complex structures

Inhibitor complex	No. of hydrogen bonds	No. of interface residues[a]	Interface accessible surface area (Å^2)	P2′ residue	$\Delta\Delta G$[b]
APPI-bt	12	21 (2)	566	Met17I	+14.4
BPTI-bt	14	20 (3)	693	Arg17I	+37.9
SBTI-pt	11	25 (4)	804	Arg65I	+27.8

[a] Residues which have any atom closer than 4 Å to an atom of the inhibitor are defined as an interface residue. The number in parenthesis is the number of substituted residues of the interface area in human trypsin IV.
[b] $\Delta G_{bt} - \Delta G_{ht\ IV}$.

Our second choice of inhibitor was BPTI since it has the same fold as APPI. Kinetic studies (Table 2) gave a K_I value for this inhibitor in complex with human trypsin IV as 1.2×10^{-6} M, almost seven orders of magnitude lower affinity than that of the bovine cationic trypsin:BPTI complex. The X-ray structure of the bovine cationic trypsin:BPTI complex showed that 20 residues of the enzyme are in closer contact than 4 Å with the inhibitor. Of these 20 residues, three are substituted in human trypsin IV (Tyr39 → Ser; Asn97 → Asp; Gly193 → Arg). Residue 97, an asparagine in bovine trypsin but an aspartate in human trypsin IV, might make favourable contacts with Arg39 of the inhibitor. This potential stabilization, however, is more than counteracted by the Gly193 → Arg substitution, which introduces a steric clash with Arg17I of BPTI in the S2′ pocket. This clash would interfere with inhibitor binding both sterically and electrostatically, and must be energetically expensive. These conclusions drawn from the model are entirely consistent with the observation that human trypsin IV binds APPI 40 times more strongly than BPTI, as determined by the ratio of their K_I values (Table 2).

A final study concerned an unrelated inhibitor, SBTI, which is approximately four times larger than APPI or BPTI. As with BPTI, this inhibitor introduces an arginine at the P2′ site. The measured K_I value for this inhibitor (Table 2) in complex with human trypsin IV was 4.2×10^{-7} M, approximately 4000 times lower than its affinity to bovine cationic trypsin. Since no structure of bovine cationic trypsin in complex with SBTI is available, structural comparisons were based upon the complex with porcine pancreatic trypsin[15] (Protein Data Bank, entry 1avw), which showed that 25 residues of porcine pancreatic trypsin form close contacts (less than 4 Å) with the bound inhibitor. Four of these 25 residues are substituted when comparing with human trypsin IV (Phe94 → Tyr (tyrosine for bovine cationic trypsin), Gly96 → Arg (serine for bovine cationic trypsin), Asn97 → Asp (asparagine for bovine cationic trypsin); Gly193 → Arg). Again it appears that the Gly193 → Arg substitution plays the dominant role in lowering the binding affinity of SBTI with human trypsin IV relative to that of bovine cationic trypsin. As with the human trypsin IV:BPTI complex, the specific conformation of Arg193 introduces both a steric and an electrostatic clash with Arg65 of SBTI.

Discussion

The X-ray structure of human trypsin IV in complex with benzamidine, in combination with comparative structural studies of other trypsin:inhibitor complexes, supports the proposal that the reduced affinity towards natural polypeptide inhibitors derives from the replacement of glycine 193 by arginine in the human enzyme.[6] In all cases modeling revealed that the positively charged side-chain of Arg193 of human trypsin IV would exert a steric clash with the inhibitor's P2′ side-chain, while in the case of SBTI and BPTI electrostatic repulsion would also occur. The important role of Arg193 in destabilising different enzyme-inhibitor complexes has recently been verified by site-directed mutagenesis studies. An Arg193 → Gly mutation of human trypsin IV decreased the values of inhibition constants by around two to three orders of magnitude, while the introduction of the Gly193 → Arg mutation into rat trypsin or human trypsin I resulted in a similar increase of the K_I values (P. Medveczky *et al*., unpublished results).

Among the polypeptide inhibitors studied here, APPI inhibits the human enzyme most strongly. In addition, the observed drop in binding affinity relative to bovine cationic trypsin was also the smallest for this natural inhibitor. This high binding affinity of APPI suggests a specific interaction, which may play a physiological role in controlling the activity of human trypsin IV in the brain. It is also significant that BPTI, which shares the same fold as APPI, binds to human trypsin IV with a K_I seven orders of magnitude greater than its binding to bovine cationic trypsin (Table 2). This equates to an estimated binding energy for BPTI with human trypsin IV only half that of its binding energy with bovine cationic trypsin (Table 2), for which it is a natural inhibitor. The rationale for this discrepancy lies in a comparison of charges along the two polypeptide inhibitors, which show that APPI has two additional negatively charged residues relative to BPTI strategically placed for binding. In addition all positively charged residues, except for the P1 arginine or lysine, which extends into the S1 binding pocket, are excluded from the interface region of the modeled APPI:human trypsin IV complex. Being anionic, APPI can form stronger interactions with human trypsin IV, which has two positive clusters of surface charges near the active site (Figure 4(a)).

The specific functional role of human trypsin IV in the brain is still unknown. In this context it should be recalled that trypsins often serve as selective activators of other protease precursors. For example, pancreatic trypsins activate with limited proteolysis proelastase and chymotrypsinogen in the pancreatic juice. Should human trypsin IV fulfill a similar role in the human brain then its proteolytic activity must be very specific, since the non-selective cleavage of proteins within the brain would potentially be disastrous. From the structural and inhibitor binding studies reported here it may be expected that the Gly193 → Arg substitution, unique to human trypsin IV, would further restrict the selection of substrates for the proteinase. Direct evidence for an enhanced substrate restriction has been provided by the observations that human trypsin IV did not activate human chymotrypsinogen,[7] nor did it activate either cationic or anionic human trypsinogen.[11] All of these zymogens are activated by both bovine cationic trypsin and human trypsin I.[11,17] Further studies should clarify if this restriction also applies to other serine proteinase zymogens (e.g. of plasminogen, prothrombin, kallikreinogen) expressed in brain. Our structural results suggest that the charge and size of P2′ residue would be critical in determining the recognition of the cleavable bond of any zymogen potentially activated by human trypsin IV.

In this context it is highly suggestive that in transgenic mice expressing human trypsin IV the overproduction of β-amyloid peptides was observed.[5] β-Amyloid peptides are produced *via* a proteolytic cleavage from Alzheimer precursor protein, which is catalysed by either β-secretase or γ-secretase.[18] Furthermore, APPI, which was shown to have the highest binding affinity to human trypsin IV of the natural inhibitors to be assayed, is also a domain of Alzheimer precursor protein. These observations, when taken together, provide circumstantial evidence supporting a hypothesis where the precursors of β-secretase are activated by human trypsin IV *via* a still unknown proteolytic cascade. The binding of the APPI domain of soluble Alzheimer precursor protein (nexin II)[19] might moderate such a pathway.

Materials and Methods

Cloning and expression

Due to the complete sequence identity of mesotrypsin and human trypsin IV we could amplify the coding sequence from a commercial human pancreatic cDNA library (Stratagen, no. 937208) with the following oligonucleotide primers: HuIV_5′: GCT GAA GCT TTC CCC GTT GAC GAT GAT GAC, HuIV_3′: GAC TGC AGA GCT CCC GGG GGC TTT AGC. These primers introduce a *Hin*dIII site at the 5′ end and a *Sac*I site at the 3′ end of the product, allowing its facile subcloning into the *Escherichia coli* expression vector pTRAP.[20] The sequence of the isolated fragments and the coding region in expression vector was verified by automated dideoxy sequencing (ABI Prism) using the Dig Dye Terminator Kit.

The yield of the expression driven by this construction was less then 1 mg/l culture. To increase the yield of the expression we transferred the coding sequence into a modified pET17b vector from which the T7 flag between restriction sites *Nde*I and *Hin*dIII was replaced by an adaptor composed of oligonucleotides AGCTTGGGTC-GAC and ATATGTCGACCCA. This modification changed the reading frame at the *Hin*dIII site. We inserted the 1037 bp coding sequence of the zymogen form of human trypsin IV, isolated after digestion with enzymes *Hin*dIII and *Sal*I into the modified pET 17b vector cut with *Hin*dIII and *Xho*I. The human trypsinogen IV was expressed and refolded as described previously.[11]

Purification

Because the crystallization experiments were sensitive to impurities the purification procedure described previously[11] had to be slightly modified. The refolded zymogen was loaded to SP-Sepharose FastFlow (Pharmacia LKB, Sweden) equilibrated in 25 mM LiOAc/HCl (pH 4.0). The column eluted in the same buffer with a linear gradient of 0 M to 1 M NaCl. Fractions containing human trypsinogen IV were pooled and dialysed against buffer A (50 mM ε-amino-caproic acid (pH 6.0), 10 mM $CaCl_2$) containing 150 mM NaCl. Human trypsinogen IV was activated by adding 0.1 ml 0.5 mg/ml bovine enterokinase, and 1.5 ml wet gel of benzamidine-Sepharose (Pharmacia LKB, Sweden) in a 50 ml Falcon tube. The benzamidine-Sepharose gel was then packed in a disposable polypropylene column (volume 5 ml) and washed with buffer A containing 350 mM NaCl. The column was eluted with 10 mM HCl, and 1 ml fractions were collected. The sample was then diluted 50× in buffer B (25 mM Tris (pH 9.0), 1 mM benzamidine/HCl), and loaded on MonoQ HR10/10 column (Pharmacia LKB, Sweden). The column was washed with 50 ml buffer B, and eluted with a linear gradient of 0 M to 1 M NaCl. The fractions containing human trypsin IV were pooled and concentrated to 0.15 ml by ultrafiltration using Centricon-10 concentrators (AMICON corp., Beverly, MA, USA). The concentrate was loaded on a Superdex 200 gel filtration column (Pharmacia LKB, Sweden) equilibrated in 0.5 M NaCl, 10 mM HCl. Fractions containing human trypsin IV were pooled and concentrated to 6.5 mg/ml. The buffer was changed to 10 mM HCl by adding 1 ml 10 mM HCl to the concentrate and repeating the concentration step one more time.

Activity measurements and inhibitor binding studies

Activity was measured on Cbz-Val-Gly-Arg-pNA substrate in a buffer containing 50 mM Tricine (pH 8.0), 0.1 M NaCl, 10 mM $CaCl_2$, 0.005 % Triton-X. The measurements were carried out in a Shimadzu UV-2101PC spectrophotometer at 37 °C. Progress curves were obtained by recording the absorbance at 410 nm for two minutes. The analysis of the progress curves was performed by the KINSIM/FITSIM program package.[21,22] Four progress curves of the uninhibited enzyme were recorded with enzyme concentrations of 3.2 nM and 6.3 nM and with substrate concentrations between 25 μM and 100 μM. The catalytic constants were determined using all progress curves in the same fitting session. The inhibited reactions were performed with 6.3 nM trypsin, 10-1000-fold molar excess of inhibitor depending on the strength of interaction and 50 μM substrate. The inhibitory constants were calculated assuming competitive inhibition using the fixed catalytic constants determined from the uninhibited reactions.

Crystallisation

Crystals were grown by the hanging drop vapour diffusion method. Equal amount of protein solution (6.5 mg/ml protein in 10 mM HCl) and precipitant solution (4 % (w/v) PEG 8000, 0.1 M Tris (pH 7.7), 5 mM $CaCl_2$, 5 mg/ml benzamidine/HCl) were mixed and equilibrated against 0.5 ml precipitant solution. Elongated crystals of about 0.2 mm × 0.3 mm × 0.6 mm were grown in one week at room temperature.

Data collection and processing

X-ray data were collected on a MAR CCD detector at BM-14 of the ESRF (λ=0.886 Å) on a crystal that was mounted in a glass capillary. The oscillation range per image was 0.25 degrees. To minimize the effect of radiation damage, the beam position on the crystal was changed several times by translation of the crystal. Data were indexed, integrated and scaled using the HKL program package.[23] The crystals belong to space group $P4_12_12$ with unit cell dimensions of $a = b = 56.68$ Å, and $c = 143.21$ Å. Subsequent data processing was carried out using programs in the CCP4 program suite.[24] Data statistics are summarized in Table 1.

Molecular replacement

The structure was solved by molecular replacement using the program AmoRe.[25] The starting coordinates were taken from the human trypsin I structure and included the protein moiety only (Protein Data Bank, entry 1trn). The rotational search in the resolution range 8-3 Å resulted in a single peak far above the noise level. The translational search assuming space group $P4_12_12$ produced the highest peak. After rigid body refinement in AMoRe the correlation coefficient was 0.66 with an R-factor of 0.37.

Model building and refinement

The model was systematically improved through iterative cycles of crystallographic refinement using the program CNS[26], and manual rebuilding by means of O.[27] The model was checked by cross-validated SigmaA weighted electron density maps calculated with both $2mF_o - DF_c$ and $mF_o - DF_c$ coefficients.[28] All refinements were performed using an amplitude-based maximum-likelihood target. A random selection of approximately 5 % of the data (test set) was assigned for calculation of the free R factor,[29] and was not included in the refinement. Residues 27, 28, 122, 128, 192 and 224 were modeled in multiple conformations.

Quality check and analysis of the final model

The stereochemistry of the model was checked using PROCHECK.[30] Least square fitting of C^α atoms was performed using LSQMAN.[31] Structural comparison with other trypsin structures was performed using Swiss-PdbViewer.[32] Existing protein-inhibitor complexes were analysed using CNS and the Protein-Protein Interaction Server.[33] Electrostatic potentials and molecular surfaces were calculated using Grasp.[34] The charges were assigned as described previously,[9] with the addition of charging the oxygen atoms of the phosphoryl group of human trypsin I (carrying -0.33 charge each).

Protein Data Bank accession code

Molecular coordinates and structure factors are available through the Protein Data Bank[35] at RCSB (accession code 1h4w).

Acknowledgements

We thank E. Kénesi and K. Béres for their contributions to this project. Dr Richard Neutze and Dr Anke Terwisscha van Scheltinga are gratefully acknowledged for careful reading of the manuscript. G.K. is supported by scholarships from ERASMUS, the Hungarian Eötvös State Fellowship and the European commission developing human potential programme. This work was supported by the Swedish Research Councils and the Hungarian Ministry of Education (FKFP).

References

1. Déry, O., Corvera, C. U., Steinhoff, M. & Bunnett, N. W. (1998). Proteinase-activated receptors: novel mechanisms of signaling by serine proteases. *Am. J. Physiol.* **274**, C1429-C1452.
2. Kato, Y., Nagashima, Y., Koshikawa, N., Miyagi, Y., Yasumitsu, H. & Miyazaki, K. (1998). Production of trypsins by human gastric cancer cells correlates with their malignant phenotype. *Eur. J. Cancer,* **34**, 1117-1123.
3. Roach, J. C., Wang, K., Gan, L. & Hood, L. (1997). The molecular evolution of the vertebrate trypsinogens. *J. Mol. Evol.* **45**, 640-652.
4. Wiegand, U., Corbach, S., Minn, A., Kang, J. & Müller-Hill, B. (1993). Cloning of the cDNA encoding human brain trypsinogen and characterization of its product. *Gene,* **136**, 167-175.
5. Minn, A., Schubert, M., Neiss, W. F. & Müller-Hill, B. (1998). Enhanced GFAP expression in astrocytes of transgenic mice expressing the human brain-specific trypsinogen IV. *Glia,* **22**, 338-347.
6. Nyaruhucha, C. N., Kito, M. & Fukuoka, S. I. (1997). Identification and expression of the cDNA-encoding human mesotrypsin(ogen), an isoform of trypsin with inhibitor resistance. *J. Biol. Chem.* **272**, 10573-10578.
7. Rinderknecht, H., Renner, I. G., Abramson, S. B. & Carmack, C. (1984). Mesotrypsin: a new inhibitor-resistant protease from a zymogen in human pancreatic tissue and fluid. *Gastroenterology,* **86**, 681-692.
8. Bode, W. & Schwager, P. (1975). The refined crystal structure of bovine beta-trypsin at 1.8 Å resolution. II. Crystallographic refinement, calcium binding site, benzamidine binding site and active site at pH 7.0. *J. Mol. Biol.* **98**, 693-717.
9. Gaboriaud, C., Serre, L., Guy-Crotte, O., Forest, E. & Fontecilla-Camps, J. C. (1996). Crystal structure of human trypsin 1: unexpected phosphorylation of Tyr151. *J. Mol. Biol.* **259**, 995-1010.
10. Gorry, M. C., Gabbaizedeh, D., Furey, W., Gates, L. K., Jr, Preston, R. A., Aston, C. E. *et al.* (1997). Mutations in the cationic trypsinogen gene are associated with recurrent acute and chronic pancreatitis. *Gastroenterology,* **113**, 1063-1068.
11. Szilágyi, L., Kénesi, E., Katona, G., Kaslik, G., Juhász, G. & Gráf, L. (2001). Comparative in vitro studies on native and recombinant human cationic trypsins. cathepsin b is a possible pathological activator of trypsinogen in pancreatitis. *J. Biol. Chem.* **276**, 24574-24580.
12. Sahin-Tóth, M. (2000). Human cationic trypsinogen. Role of Asn-21 in zymogen activation and implications in hereditary pancreatitis. *J. Biol. Chem.* **275**, 22750-22755.
13. Scheidig, A. J., Hynes, T. R., Pelletier, L. A., Wells, J. A. & Kossiakoff, A. A. (1997). Crystal structures of bovine chymotrypsin and trypsin complexed to the inhibitor domain of Alzheimer's amyloid beta-protein precursor (APPI) and basic pancreatic trypsin inhibitor (BPTI): engineering of inhibitors with altered specificities. *Protein Sci.* **6**, 1806-1824.
14. Helland, R., Otlewski, J., Sundheim, O., Dadlez, M. & Smålas, A. O. (1999). The crystal structures of the complexes between bovine beta-trypsin and ten P1 variants of BPTI. *J. Mol. Biol.* **287**, 923-942.
15. Song, H. K. & Suh, S. W. (1998). Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight

into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. *J. Mol. Biol.* **275**, 347-363.
16. Walter, J., Steigemann, W., Singh, T. P., Bartunik, H., Bode, W. & Huber, R. (1982). On the disordered activation domain in trypsinogen. chemical labelling and low-temperature crystallography. *Acta Crystallog. sect. B*, **38**, 1462-1472.
17. Rinderknecht, H. (1986). Activation of pancreatic zymogens. Normal activation, premature intrapancreatic activation, protective mechanisms against inappropriate activation. *Dig. Dis. Sci.* **31**, 314-321.
18. Kang, J., Lemaire, H. G., Unterbeck, A., Salbaum, J. M., Masters, C. L., Grzeschik, K. H. *et al*. (1987). The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. *Nature*, **325**, 733-736.
19. Oltersdorf, T., Fritz, L. C., Schenk, D. B., Lieberburg, I., Johnson-Wood, K. L., Beattie, E. C. *et al*. (1989). The secreted form of the Alzheimer's amyloid precursor protein with the Kunitz domain is protease nexin-II. *Nature*, **341**, 144-147.
20. Gráf, L., Craik, C. S., Patthy, A., Roczniak, S., Fletterick, R. J. & Rutter, W. J. (1987). Selective alteration of substrate specificity by replacement of aspartic acid-189 with lysine in the binding pocket of trypsin. *Biochemistry*, **26**, 2616-2623.
21. Zimmerle, C. T. & Frieden, C. (1989). Analysis of progress curves by simulations generated by numerical integration. *Biochem. J.* **258**, 381-387.
22. Dang, Q. & Frieden, C. (1997). New PC versions of the kinetic-simulation and fitting programs, KINSIM and FITSIM. *Trends Biochem. Sci.* **22**, 317.
23. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
24. Bailey, S. (1994). The CCP4 Suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760-763.
25. Navaza, J. & Saludjian, P. (1997). AMoRe: an automated molecular replacement program package. *Methods Enzymol.* **276**, 581-594.
26. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al*. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905-921.
27. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110-119.
28. Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallog. sect. A*, **42**, 140-149.
29. Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472-475.
30. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.
31. Kleywegt, G. J. (1999). Experimental assessment of differences between related protein crystal structures. *Acta Crystallog. sect. D*, **55**, 1878-1884.
32. Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714-2723.
33. Jones, S. & Thornton, J. M. (1996). Principles of protein-protein interactions. *Proc. Natl Acad. Sci. USA*, **93**, 13-20.
34. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281-296.
35. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al*. (2000). The Protein Data Bank. *Nucl. Acids Res.* **28**, 235-242.
36. Castro, M. J. & Anderson, S. (1996). Alanine point-mutations in the reactive region of bovine pancreatic trypsin inhibitor: effects on the kinetics and thermodynamics of binding to beta-trypsin and alpha-chymotrypsin. *Biochemistry*, **35**, 11435-11446.
37. Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H. & Blow, D. M. (1974). Crystal structure of the complex of porcine trypsin with soybean trypsin inhibitor (Kunitz) at 2.6-Å resolution. *Biochemistry*, **13**, 4212-4228.
38. Kitaguchi, N., Takahashi, Y., Oishi, K., Shiojiri, S., Tokushima, 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.
39. Shotton, D. M. & Hartley, B. S. (1970). Amino-acid sequence of porcine pancreatic elastase and its homologies with other serine proteinases. *Nature*, **225**, 802-806.

Edited by R. Huber

(Received 16 August 2001; received in revised form 27 November 2001; accepted 4 December 2001)