Protein Science (1995), *4*:960-972. Cambridge University Press. Printed in the USA.

Crystal structure of human pepsin and its complex with pepstatin

MASAO FUJINAGA,[1] MAIA M. CHERNAIA,[1] NADYA I. TARASOVA,[2] STEVE C. MOSIMANN,[1] AND MICHAEL N.G. JAMES[1]

[1] Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada
[2] Molecular Aspects of Drug Design Section, ABL-Basic Research Program, National Cancer Institute, FCDRC, P.O. Box B, Frederick, Maryland 21702

(RECEIVED January 26, 1995; ACCEPTED February 24, 1995)

Abstract

The three-dimensional crystal structure of human pepsin and that of its complex with pepstatin have been solved by X-ray crystallographic methods. The native pepsin structure has been refined with data collected to 2.2 Å resolution to an *R*-factor of 19.7%. The pepsin:pepstatin structure has been refined with data to 2.0 Å resolution to an *R*-factor of 18.5%. The hydrogen bonding interactions and the conformation adopted by pepstatin are very similar to those found in complexes of pepstatin with other aspartic proteinases. The enzyme undergoes a conformational change upon inhibitor binding to enclose the inhibitor more tightly. The analysis of the binding sites indicates that they form an extended tube without distinct binding pockets. By comparing the residues on the binding surface with those of the other human aspartic proteinases, it has been possible to rationalize some of the experimental data concerning the different specificities. At the S1 site, valine at position 120 in renin instead of isoleucine, as in the other enzymes, allows for binding of larger hydrophobic residues. The possibility of multiple conformations for the P2 residue makes the analysis of the S2 site difficult. However, it is possible to see that the specific interactions that renin makes with histidine at P2 would not be possible in the case of the other enzymes. At the S3 site, the smaller volume that is accessible in pepsin compared to the other enzymes is consistent with its preference for smaller residues at the P3 position.

Keywords: aspartic proteinase; binding site; crystal structure; inhibitor

Pepsin, the well-known aspartic proteinase, is produced by the human gastric mucosa in seven different zymogen isoforms (Samloff, 1969; Foltmann, 1981). These have been subdivided into two types: pepsinogen I (pepsinogen A), consisting of PGA 1-5, and pepsinogen II (pepsinogen C or progastricsin), consisting of PGC 6 and 7. Three major species of pepsinogen A—PGA 3, 4, 5—have been sequenced; these sequences show that PGA 3 and 5 differ only in the propart so that after conversion to the mature enzymes, the resulting pepsins are identical (Sogawa et al., 1983; Evers et al., 1988, 1989).

Among the other aspartic proteinases produced by human tissue are renin, cathepsin D, and cathepsin E. Renin is a highly specific enzyme involved in the regulation of blood pressure and sodium and volume homeostasis. It cleaves angiotensinogen to produce angiotensin I, which is subsequently cleaved by the angiotensin converting enzyme (ACE) to produce angiotensin II. Angiotensin II is one of the most potent vasoconstrictors known. It also stimulates the release of aldosterone resulting in sodium and water retention (Cody, 1994). Cathepsin D is a widespread lysosomal enzyme involved in protein catabolism. It has also been implicated in disease states such as Alzheimer β-amyloid formation (Cataldo & Nixon, 1990) and metastasis of breast cancers (Rochefort, 1990). Cathepsin E corresponds to the electrophoretically slow-moving protease found in gastric mucosa and has been found as well in the erythrocyte membrane (Tarasova et al., 1986). Its function has not been well characterized but recently it has been shown to be responsible for some antigen processing (Bennett et al., 1992) and it may also have a role in the production of the vasoconstrictor, endothelin (Lees et al., 1990).

The structural study of human pepsin and other aspartic proteinases has great applicability in drug design. To insure oral bioavailability and specific action of drugs aimed at inhibiting aspartic proteinases such as renin or HIV protease, the compound must not bind to pepsin or other aspartic proteinases in the body. To this end, it is important to understand the origin of subsite specificities and the differences among the specificities of the human enzymes. We have determined the crystal structures of human pepsin alone and in complex with an aspar-

Reprint requests to: Michael N.G. James, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada; e-mail: miik@biochem.ualberta.ca.

tic proteinase inhibitor, pepstatin. The analyses of these structures have given insights into the different specificities of human aspartic proteinases.

Results

Refinement results

The refined structure for the native human pepsin consists of 2,438 protein atoms and 102 water molecules. The *R*-factor is 19.7% for the 21,303 reflections between 30 and 2.2 Å resolution. The occupancies for the residues Val 1, Ser 241, Asp 242, Gly 243, Ser 295, and Gly 296 have been set to zero because there is insufficient electron density to assign their positions. The human pepsin:pepstatin complex consists of 2,438 protein atoms, 48 inhibitor atoms, and 308 water molecules. The final *R*-factor is 18.5% for the 30,584 reflections between 30 and 2.0 Å resolution. The occupancies of residues Val 1, Glu 239, Asn 240, Ser 241, Asp 242, Gly 243, and Glu 294 were also set to zero. These atoms with zero occupancy still affect the behavior of the refinement through their stereochemical restraints with the rest of the molecule. However, the lack of information from the X-ray diffraction data often results in unfavorable conformations for these regions. The positions of the atoms with zero occupancies are considered to be indeterminate.

Quality of the models

The qualities of the structural models were assessed by the PROCHECK suite of programs (Laskowski et al., 1993). Figure 1 shows the ϕ–ψ plots. In both cases, 90% of the residues fall within the most favored regions. There is only one residue, Glu 279, in native pepsin that lies in the generously allowed region. This residue has high *B*-factors and its position is not well determined. Tables 1, 2, and 3 summarize the mean and standard deviations of stereochemical parameters, bond lengths, and bond angles, respectively.

The overall error in the coordinates was estimated using σ_a plots (Read, 1986) as shown in Figure 2. They indicate an RMS error of 0.27 Å for native pepsin and 0.23 Å for pepsin:pepstatin complex.

The reliability of the structure determination along the polypeptide chain can be obtained from the real space fit (Jones et al., 1991; Jones & Kjeldgaard, 1993) or the *B*-factor. Figure 3 shows the two plots superposed for each structure. The correlation between the two representations is evident. This may be expected because both are measures of the amount of electron density around the atomic position. Most of the poorly fitting regions correspond to loops in the protein that tend to be more mobile. In particular, the high degree of disorder and the resultant lack of reliability of the regions around residues 242 and 295 are clearly seen.

As an example of the quality of the model, the electron density corresponding to the region of the pepstatin molecule is shown in Figure 4. The central part of the molecule is well defined, but the density is rather poor for the Iva residue and the terminal Sta residue. These residues must adopt multiple conformations because the fit of the model cannot be made convincing. These residues have average *B*-factors of 42 and 64 Å^2, respectively.

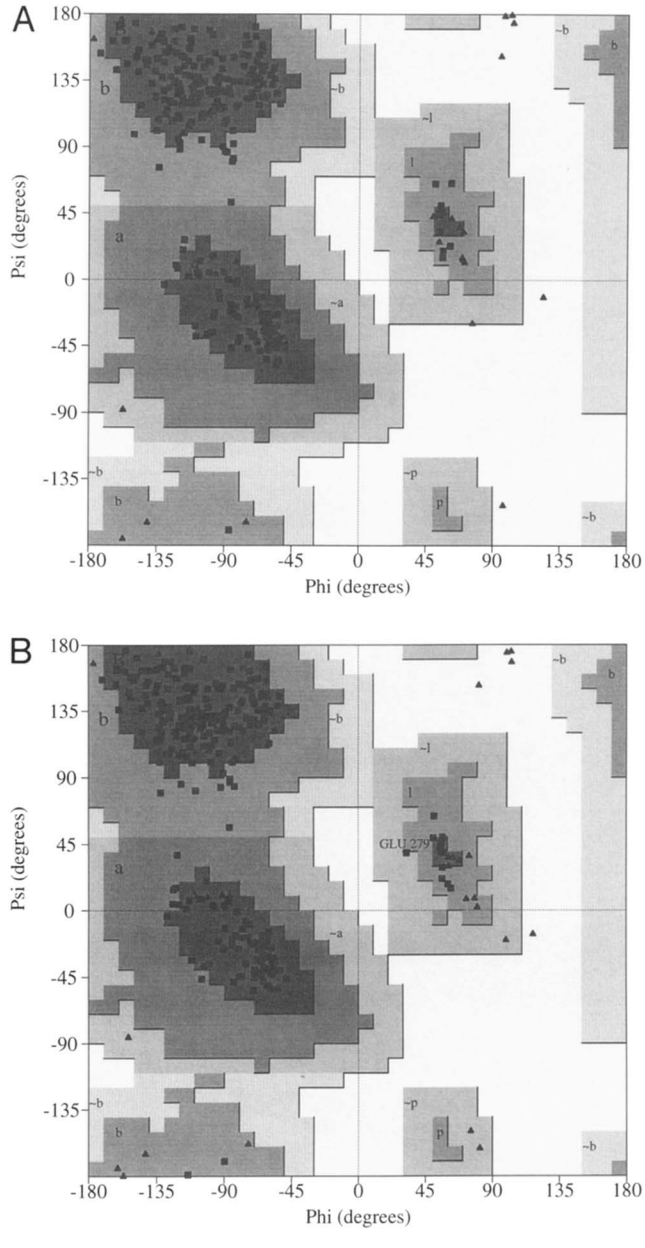

Fig. 1. ϕ–ψ plots as produced by the PROCHECK program (Laskowski et al., 1993). Glycine residues are shown as triangles, all the others as squares. The different regions, in decreasing darkness, correspond to most favored regions, additional allowed regions, generously allowed regions, and disallowed regions. **A:** Native pepsin. **B:** Pepsin:pepstatin complex.

Overall description

The overall fold of the aspartic proteinases has been well documented (Bott et al., 1982; James & Sielecki, 1983; Blundell et al., 1985). In particular, the structure of human pepsin follows closely the structure of porcine pepsin described previously (Sielecki et al., 1990). As was done for porcine pepsin, the structure will be divided into three domains (see Kinemage 1). The central domain consists of a six-stranded antiparallel β-sheet that serves as a backbone to the active-site region of the molecule. It is made up of residues Val 1–Leu 6, Asp 149–Val 184, and

Table 1. *Statistics of stereochemical parameters*

Stereochemical parameter	Ideal[a] mean (SD)	Native pepsin mean (SD)	Pepstatin:pepsin complex mean (SD)
ϕ-ψ in most favored regions of Ramachandran plot	>90%	90.2%	89.8%
χ_1 Dihedral angle (°)			
Gauche minus	64.1 (15.7)	61.5 (10.3)	63.7 (9.4)
Trans	183.6 (16.8)	184.8 (12.2)	184.4 (11.3)
Gauche plus	−66.7 (15.0)	−64.4 (10.9)	−64.6 (11.5)
χ_2 Dihedral angle (°)	177.4 (18.5)	171.5 (14.8)	173.5 (13.8)
Proline ϕ torsion angle (°)	−65.4 (11.2)	−66.8 (7.4)	−67.2 (8.8)
Helix ϕ torsion angle (°)	−65.3 (11.9)	−65.1 (12.2)	−66.5 (11.2)
Helix ψ torsion angle (°)	−39.4 (11.3)	−37.1 (11.9)	−36.4 (11.8)
χ_3 (S-S bridge) (°)			
Right-handed	96.8 (14.8)	92.9 (22.1)	89.0 (17.7)
Left-handed	−85.8 (10.7)	−96.2 (0.0)	−95.1 (0.0)
Disulfide bond separation (Å)	2.0 (0.1)	2.0 (0.0)	2.0 (0.0)
ω Dihedral angle (°)	180 (5.8)	177.7 (5.6)	177.8 (6.2)
Main-chain hydrogen bond energy (kcal/mol)[b]	−2.03 (0.75)	−2.1 (0.8)	−2.1 (0.7)
Cα chirality: ζ "virtual" torsion angle (°) (Cα-N-C-Cβ)	33.9 (3.5)	33.9 (1.8)	34.0 (1.8)

[a] Morris et al. (1992).
[b] As defined by Kabsch and Sander (1983).

Gln 308–Ala 326. The N-terminal lobe is composed of residues Glu 7–Gln 148 and the C-terminal lobe is made up of residues Thr 185–Arg 307. The lobes consist of orthogonally packed β-sheets, with the N- and C-terminal lobes having three and two layers, respectively. The overall structure and its division into three domains can be appreciated in Figure 5, which shows the main-chain atoms with the associated hydrogen bonds.

Interactions with pepstatin and substrate binding sites

As observed for other complexes of inhibitors bound to aspartic proteinases, pepstatin adopts an extended conformation with the first statyl hydroxyl oxygen occupying a position in the active site between the carboxyl groups of Asp 32 and Asp 215. This is a position occupied by a conserved water molecule in native structures of all aspartic proteinases. The hydrogen bonding pattern between the inhibitor and the enzyme (Fig. 6; Table 4) is also well conserved in other structurally determined complexes with pepstatin (Suguna et al., 1992; Bailey et al., 1993; Baldwin et al., 1993). The residues making van der Waals contact with the side chains of pepstatin are listed in Table 5. It should be noted that designations for P1′, P2′, and P3′ (Schechter & Berger, 1967) are only approximate because the statine residue has two carbon atoms more compared to a normal

Table 2. *Statistics of bond length (Å)*

Bond	Ideal[a] mean (SD)	Native pepsin mean (SD)	Pepstatin:pepsin complex mean (SD)
C-N			
Except Pro	1.329 (0.014)	1.328 (0.007)	1.328 (0.008)
Pro	1.341 (0.016)	1.340 (0.006)	1.342 (0.006)
C-O	1.231 (0.020)	1.228 (0.006)	1.228 (0.006)
C$^\alpha$-C			
Except Gly	1.525 (0.021)	1.535 (0.009)	1.534 (0.009)
Gly	1.516 (0.018)	1.535 (0.008)	1.531 (0.011)
C$^\alpha$-C$^\beta$			
Ala	1.521 (0.033)	1.530 (0.005)	1.530 (0.008)
Ile, Thr, Val	1.540 (0.027)	1.540 (0.007)	1.543 (0.009)
The rest	1.539 (0.020)	1.533 (0.007)	1.534 (0.008)
N-C$^\alpha$			
Except Gly, Pro	1.458 (0.019)	1.476 (0.007)	1.477 (0.008)
Gly	1.451 (0.016)	1.475 (0.007)	1.477 (0.008)
Pro	1.466 (0.015)	1.478 (0.005)	1.473 (0.005)

[a] Engh and Huber (1991).

Table 3. *Statistics of bond angles (°)*

Angle	Ideal[a] mean (SD)	Native pepsin mean (SD)	Pepstatin:pepsin complex mean (SD)
C^{α}-C-N			
Except Gly, Pro	116.2 (2.0)	115.2 (1.7)	115.7 (1.8)
Gly	116.4 (2.1)	114.9 (1.9)	115.6 (2.1)
Pro	116.9 (1.5)	115.6 (1.4)	116.4 (1.2)
O-C-N			
Except Pro	123.0 (1.6)	123.6 (1.1)	123.2 (1.2)
Pro	122.0 (1.4)	123.4 (0.7)	122.7 (1.0)
C-N-C^{α}			
Except Gly, Pro	121.7 (1.8)	123.3 (1.4)	122.9 (1.6)
Gly	120.6 (1.7)	123.2 (1.2)	123.2 (1.7)
Pro	122.6 (5.0)	122.7 (1.3)	122.7 (1.2)
C^{α}-C-O			
Except Gly	120.8 (1.7)	120.8 (1.2)	120.7 (1.5)
Gly	120.8 (2.1)	120.8 (1.2)	120.6 (1.8)
C^{β}-C^{α}-C			
Ala	110.5 (1.5)	110.0 (1.5)	109.8 (1.7)
Ile, Thr, Val	109.1 (2.2)	111.1 (2.2)	111.0 (2.2)
The rest	110.1 (1.9)	110.1 (2.1)	110.2 (2.2)
N-C^{α}-C			
Except Gly, Pro	111.2 (2.8)	111.3 (2.7)	111.0 (2.7)
Gly	112.5 (2.9)	113.0 (2.1)	112.9 (2.5)
Pro	111.8 (2.5)	112.1 (2.3)	112.0 (2.5)
N-C^{α}-C^{β}			
Ala	110.4 (1.5)	109.9 (1.1)	110.3 (1.2)
Ile, Thr, Val	111.5 (1.7)	111.5 (2.0)	111.6 (2.3)
Pro	103.0 (1.1)	105.1 (1.0)	104.9 (1.0)
The rest	110.5 (1.7)	110.4 (1.6)	110.4 (1.9)

[a] Engh and Huber (1991).

amino acid residue. Most of the interactions are made with the residues P1, P2, and P3. The relatively large number of contacts made by the P3′ Sta should be considered with less weight because this residue is disordered. The structures of pepstatin bound to cathepsin D, rhizopuspepsin, endothiapepsin, and to human pepsin have been superposed in Figure 7. It shows that the inhibitor binds in a very similar conformation to all enzymes except for the extremities.

Changes on binding

There is a conformational change of pepsin on binding pepstatin. The change is relatively small, with an RMS difference in the coordinates of all the α-carbon atoms being 0.33 Å after superposition. Nevertheless, it can be clearly seen as a relative movement of the domains to enclose the inhibitor more closely. The nature of the movement can best be appreciated as an ani-

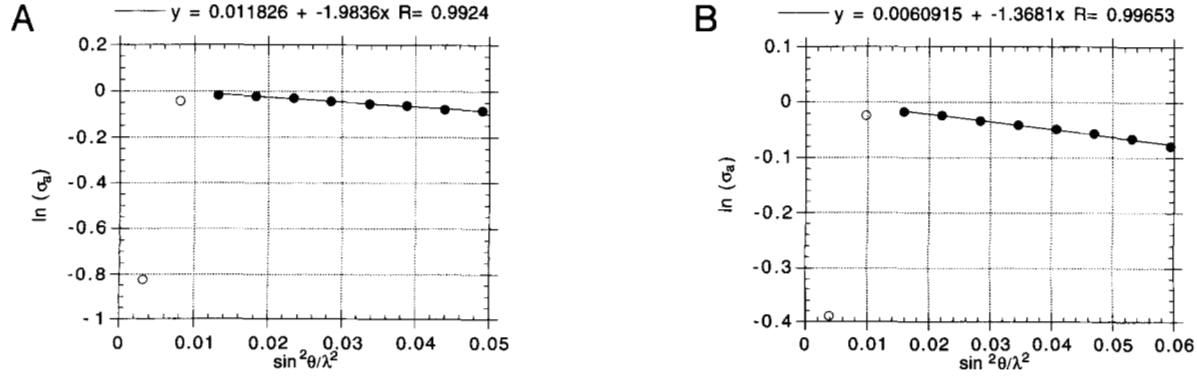

Fig. 2. Analysis of coordinate error using a σ_a plot (Read, 1986). Least-squares lines have been fitted to the solid points. **A:** Native pepsin. **B:** Pepsin:pepstatin complex.

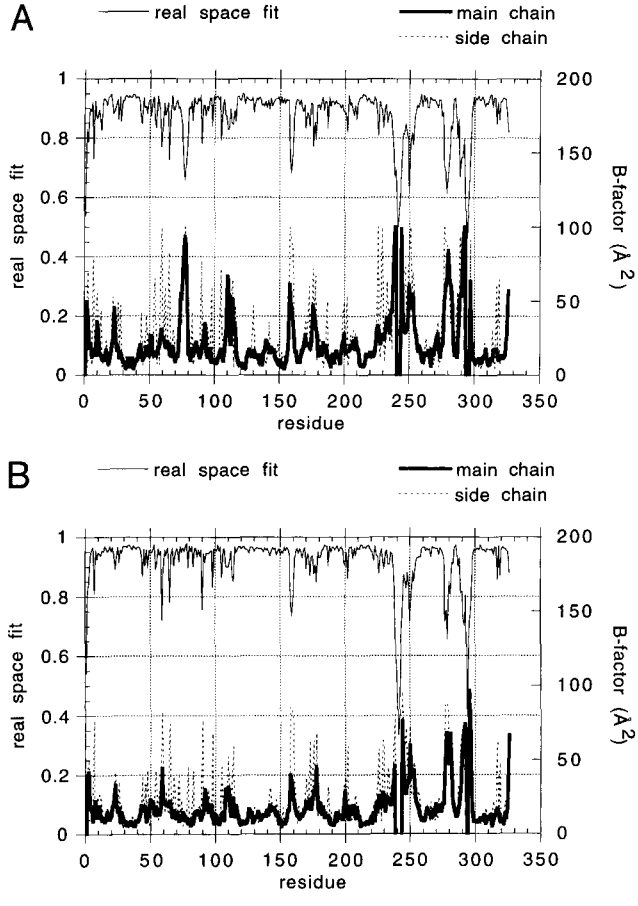

Fig. 3. Real space fit (Jones et al., 1991; Jones & Kjeldgaard, 1993) and average *B*-factor along the polypeptide chain. Residues for which the occupancies were set to zero were given a *B*-factor of zero for the purpose of this plot only. **A:** Native pepsin. **B:** Pepsin:pepstatin complex.

Table 4. *Intermolecular hydrogen bonds between pepsin and pepstatin*[a]

Pepstatin		Pepsin		Distance (Å)
Val I2	N	Ser 219	O^{γ}	2.9 (2.8)[b]
Val I2	O	Ser 219	N	3.0 (3.1)
Val I3	N	Thr 77	$O^{\gamma 1}$	3.1 (4.0)
Val I3	O	Gly 76	N	3.0 (3.5)
		Thr 77	N	3.0 (4.1)
Sta I4	N	Gly 217	O	2.9 (3.0)
Sta I4	OH	Gly 217	O	3.5 (3.9)
		Asp 32	$O^{\delta 1}$	2.6 (2.6)
		Asp 215	$O^{\delta 1}$	2.8 (2.8)
		Asp 215	$O^{\delta 2}$	3.0 (2.8)
Sta I4	O	Gly 76	N	3.0 (3.6)
Ala I5	N	Gly 34	O	2.9 (2.9)
Ala I5	O	Tyr 189	O^{η}	2.7 (2.2)
Sta I6	N	Thr 74	O	2.9 (3.3)

[a] Hydrogen bonds have been defined as those interactions between donors and acceptors with a distance less than 3.5 Å and an angle at the hydrogen atom (where the position is unambiguous) of greater than 110° and the angle at the acceptor atom greater than 80°.

[b] Numbers in parentheses are the distances that pepstatin would make if the structure of native pepsin were superimposed on the complex structure.

Table 5. *Inhibitor side-chain van der Waals contacts to pepsin*

Inhibitor	Enzyme	Total number of contacts
P4 Iva I1	Met 12	3
	Ser 219	
	Gln 287	
p3 Val I2	Met 12	8
	Thr 77	
	Phe 111	
	Gly 217	
	Ser 219	
P2 Val I3	Gly 76	4
	Thr 218	
	Gln 287	
	Met 289	
P1 Sta I4	Val 30	12
	Asp 32	
	Tyr 75	
	Thr 77	
	Ile 120	
	Gly 217	
P2′ Ala I5	Ser 25	2
	Gly 34	
P3′ Sta I6	Thr 74	6
	Tyr 75	
	Gly 76	
	Leu 291	

mation (Kinemage 1), where the correlated movements are best seen. However, such a representation suffers because what one perceives as movement is dependent on how the superposition of the two molecules is done. An unbiased analysis can be made with a difference diagonal distance matrix (Fig. 8). This plot shows the absolute values of the differences in the interatomic distances between all α-carbon atom pairs. Rigid domains are expected have small values and movements of one domain with respect to another will show up as high values. In either case, it can be seen that the C-terminal domain makes a large movement with respect to the rest of the molecule, whereas the N-terminal and the central domain move to a smaller extent with respect to each other. This larger movement or the flexibility of the C-terminal domain has been noted for other aspartic proteinases (Šali et al., 1989, 1992; Abad-Zapatero et al., 1990; Sielecki et al., 1990). In the case of penicillopepsin, the binding of an inhibitor results in an opening of the enzyme with the central and C-terminal domains moving together and away from the inhibitor. For rhizopuspepsin, no movement was reported but an analysis using animation as was done with Kinemage 1 indicates a very small movement toward the inhibitor. Figure 8 and Kinemage 1 also indicate that one of the largest changes occurs in the region of the "flap" (residues 71–82). This region also undergoes a drastic reduction in the *B*-factor upon binding the inhibitor (Fig. 3). The large number of noncovalent interactions that this loop makes with pepstatin (Tables 4, 5) presumably are responsible for the conformational change and the loss of flexibility. The different kinds of movement seen indicate that the

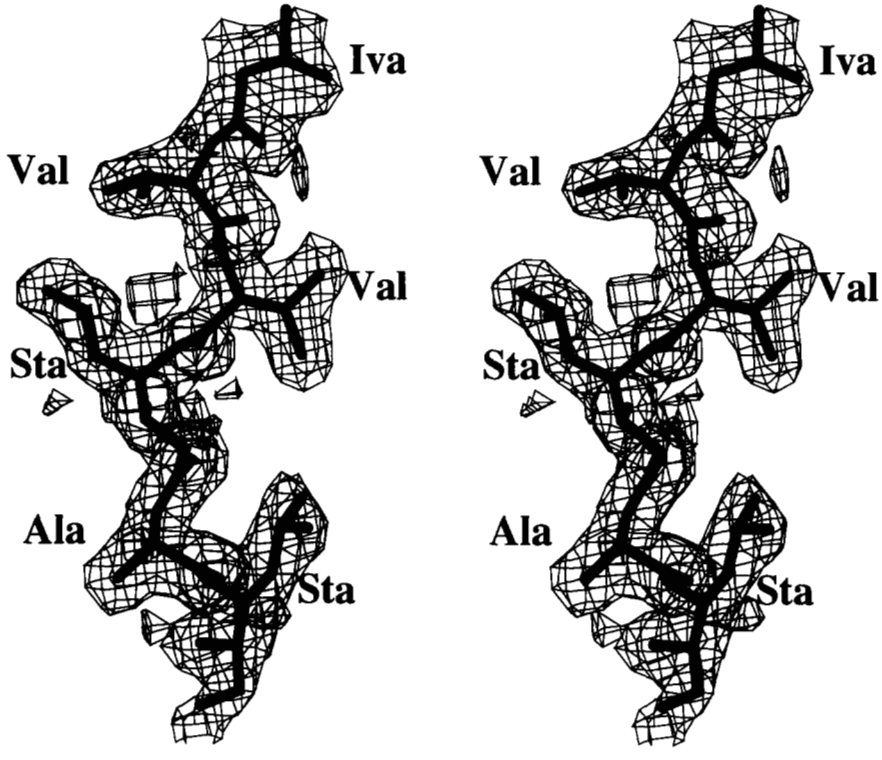

Fig. 4. Stereogram showing the pepstatin molecule and its associated electron density. Electron density was calculated with coefficients $(2m|F_o| - D|F_c|, \alpha_c)$ (Read, 1986) and contoured at 1σ.

enzyme is capable of adjusting to fit optimally to an inhibitor. These data still do not answer whether a particular conformational change is required for catalysis, as has been suggested (Fruton, 1976), because these movements are seen when an inhibitor, not a substrate, is bound.

Comparison with other human enzymes

The amino acid sequence of human pepsin is compared to those of the other human aspartic proteinases as well as to that of porcine pepsin in Figure 9. The similarities of human pepsin to the other enzymes range from 28% identity with renin to 53% with cathepsin E and 84% with porcine pepsin. The structural similarities among the enzymes for which the structures have been determined are summarized in Table 6. It shows the RMS deviations of the equivalent α-carbon atoms after superposition with the program O (Jones & Kjeldgaard, 1993). The structural similarity correlates with the similarity in the sequences. This would imply that for the comparative model building of cathepsin E, the structure of human pepsin would serve as the best starting point because this has the highest sequence identity.

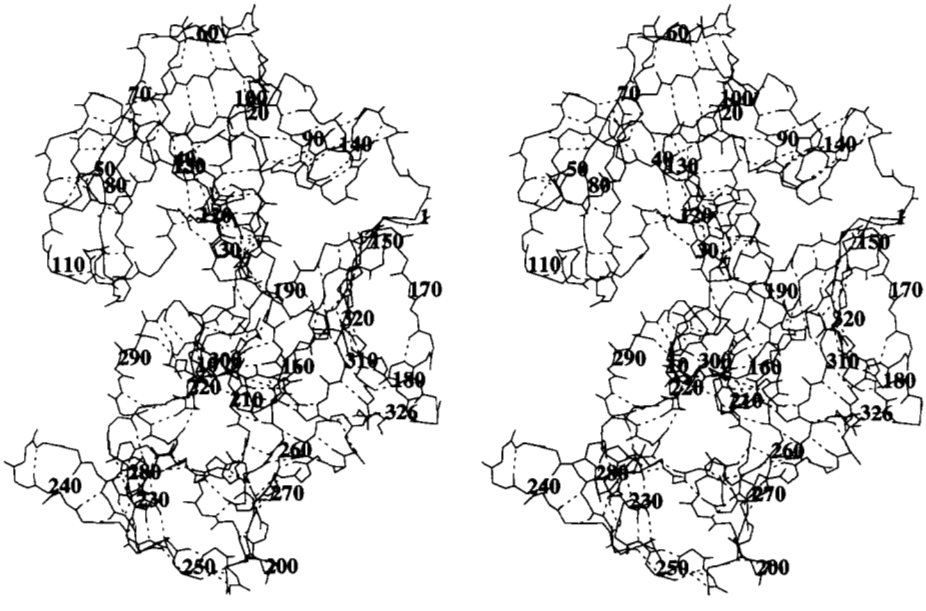

Fig. 5. Stereogram showing the overall structure of human pepsin. Main-chain atoms are shown with the associated hydrogen bonds shown with broken lines.

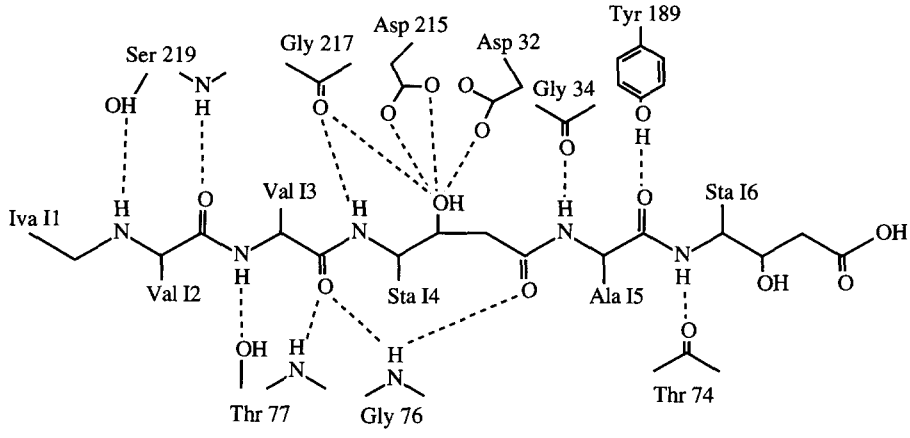

Fig. 6. Stylized representation of the potential hydrogen bonding interactions (dashed) between pepsin and pepstatin.

Discussion

Comparison of binding sites with other human enzymes

There are limited data available for the specificities of the human aspartic proteinases and many of these have been summarized in Table 7. These data have been obtained from binding studies of peptide-based inhibitors.

As a starting point in trying to rationalize the binding data in structural terms, solvent-accessible surfaces were calculated (Kleywegt & Jones, 1994) and displayed as chicken wire representations using O (Jones & Kjeldgaard, 1993) for the three human enzymes, pepsin, renin, and cathepsin D, for which the structures are known (Fig. 10). It is immediately obvious that the binding sites form a continuous tube and that separate pockets for each subsite cannot easily be discerned. In addition, what may be called sites S1 and S3 communicate with each other and likewise for sites S2 and S1′. This implies that the specificity at S3, for example, will depend on the residue at P1. The dangers of assuming independence of binding sites have been pointed out recently for the case of inhibitor binding to renin (Epps et al., 1990). Such dependence could also result if binding at one site affects the conformation of the inhibitor in such a way as to modify the binding at another site. Another consequence of the ill-defined binding sites is that a particular residue can bind in more than one conformation. This has actually been observed for the P2 residue where different conformations of a histidine residue have been observed in inhibitor complexes with endo-

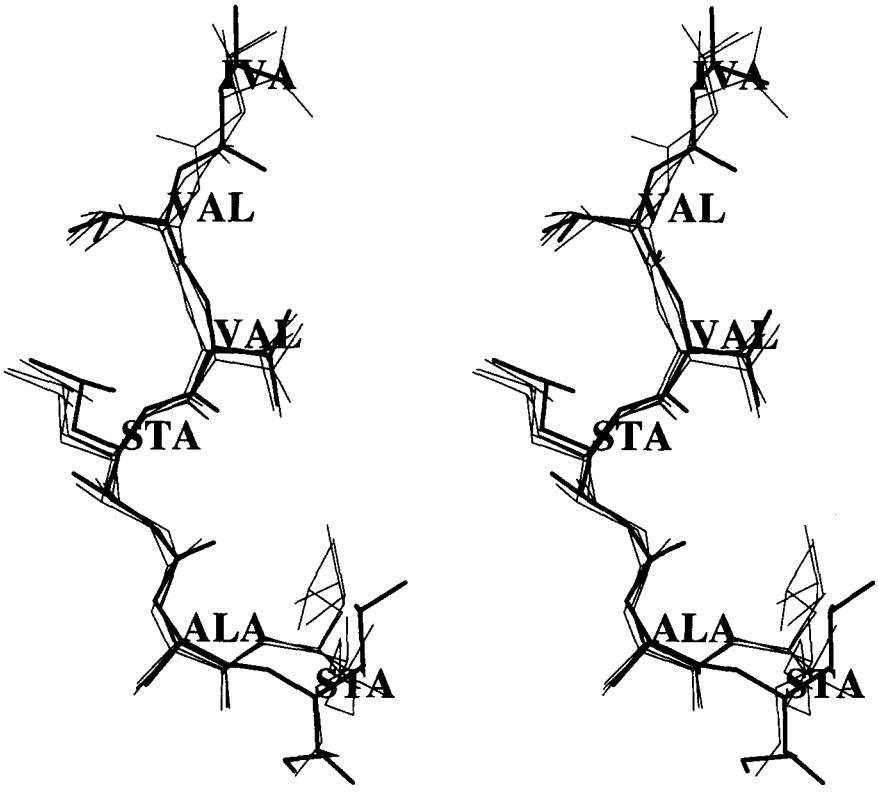

Fig. 7. Stereogram showing the superposition of structures of pepstatin bound to pepsin (thick lines), to endothiapepsin (Bailey et al., 1993), to rhizopuspepsin (Suguna et al., 1992), and to cathepsin D (Baldwin et al., 1993).

Table 6. *Structural alignments*[a]

	Pepsin	Pepsin: pepstatin	Renin (1rne)	Cathepsin D (1lyb)	Porcine pepsin (4pep)
Pepsin		0.33 (326)	1.43 (311)	1.10 (315)	0.60 (325)
Pepsin:pepstatin	100%		1.35 (311)	1.06 (321)	0.65 (325)
Renin	28%	28%		1.15 (315)	1.44 (304)
Cathepsin D	34%	34%	38%		1.16 (314)
Porcine pepsin	84%	84%	34%	41%	

[a] Upper triangle shows RMS deviations (Å) and number of α-carbon atoms superimposed in parentheses. Lower triangle gives the sequence homology in terms of percent identity.

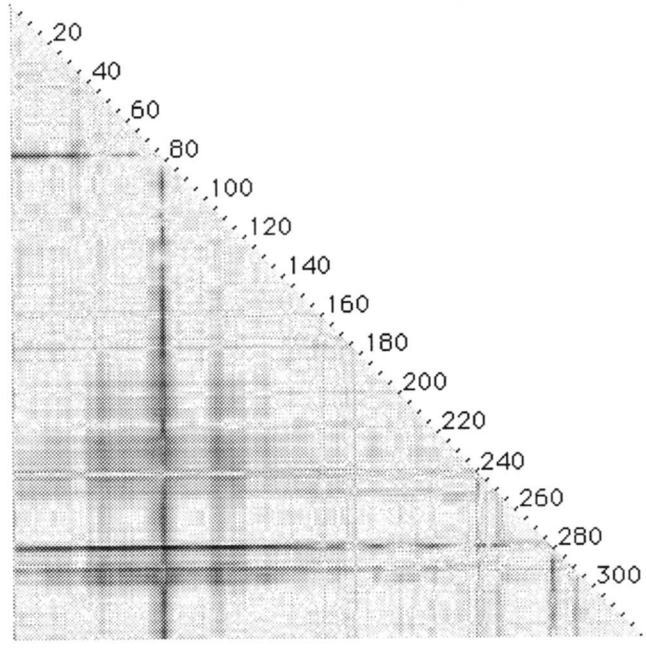

Fig. 8. Difference distance plot. Each point represents the absolute value of the difference in the distance between the ith and jth α-carbon atom coordinates in the native and the complex structures (i.e., $|d_{i,j}(\text{native}) - d_{i,j}(\text{complex})|$). The darker the points, the larger the magnitude in the difference.

thiapepsin and with rhizopuspepsin (Foundling et al., 1987; Suguna et al., 1992).

The visual inspection of the accessible surfaces also shows that pepsin and cathepsin D are very similar, with cathepsin D having a slightly larger S3. Renin, on the other hand, has a much larger S3 and smaller S1′ than either pepsin or cathepsin D.

Hydrophobic residues are preferred by all the enzymes at P1. Renin prefers large hydrophobic residues, whereas all the others except gastricsin bind smaller hydrophobic residues better. Gastricsin has a mixed specificity at P1 favoring Phe over Leu but not over a cyclohexylalanyl residue. Valine at position 120 in human renin compared to isoleucine in the other enzymes allows this enzyme to accommodate larger residues at P1. It is not clear how the mixed specificity of gastricsin arises. The only difference in this region would be the Asn at position 111 that replaces a Phe in pepsin, but it is not evident how these changes could result in the observed P1 specificity of gastricsin.

As mentioned above, the multiple conformations observed for P2 residues in several of the inhibitors make the analysis of the

Table 7. *Enzyme specificities*[a]

	P3	P2	P1
Pepsin	Ala ≫ Phe[b]	Tyr ≫ Ala ≫ His[b] Phe ≫ His[b] Val ≫ Phe > Glu ≫ Ala ≫ Lys[c]	Leu ≫ Phe > CH[b] Leu > CH[b]
Gastricsin	Phe > Ala (slight preference)[b]	Tyr = Ala ≫ His[b] Phe ≫ His[b] Glu ≫ Lys ≫ Ala > Phe ≫ Val (pH 1.7)[c] Lys > Lys ≫ Ala > Phe ≫ Val (pH 4.5)[c]	Phe > Leu > CH[b] Leu ≫ CH[b]
Renin	Phe ≫ Ala[b]	His ≫ Tyr ≫ Ala[b] His ≫ Phe[b]	CH ≫ Phe ≫ Leu[b] CH ≫ Leu[b]
Cathepsin D	Phe ≫ Ala[b] Ile > Ala[d]	His > Ala ≫ Tyr[b] His ≫ Phe[b] Leu > Glu > Ile > Ala > Val > Ser > Asp > Gln > Arg > Lys > His[d] Leu > Glu > Asp > Ala > Ser > Arg > Lys[d]	Leu > CH ≫ Phe[b] Leu ≫ CH[b]
Cathepsin E	Phe ≫ Ala[b]	Tyr = Ala ≫ His[b] Phe ≫ His[b]	CH > Leu > Phe[b] Leu ≫ CH[b]

[a] CH, cyclohexylalanyl.
[b] Jupp et al. (1990).
[c] Baxter et al. (1990).
[d] Scarborough et al. (1993).

```
pp  ----------IGDEPLENYLDTEYFGTIGIGTPAQDFTVIFDTGSSNLWVPSVYCSSL-
hp  ----------VDEQPLENYLDMEYFGTIGIGTPAQDFTVVFDTGSSNLWVPSVYCSSL-
ga  ----------SVTYEPMAYMDAAYFGEISIGTPPQNFLVLFDTGSSNLWVPSVYCQSQ-
rn  -----LTLGNTTSSVILTNYMDTQYYGEIGIGTPPQTFKVVFDTGSSNVWVPSSKCSRLY
cd  ----------GPIPEVLKNYMDAQYYGEIGIGTPPQCFTVVFDTGSSNLWVPSIHCKLLD
ce  IQFTESCSMDQSAKEPLINYLDMEYFGTISIGSPPQNFTVIFDTGSSNLWVPSVYCTSP-
                     *.*  *.* * **.* * * *.*******.****  *
                1       10        20        30        40

pp  -ACSDHNQFNPDDSSTFEATSQELSITYGTGSMTGILGYDTVQVGGISDT----------
hp  -ACTNHNRFNPEDSSTYQSTSETVSITYGTGSMTGILGYDTVQVGGISDT----------
ga  -ACTSHSRFNPSESSTYSTNGQTFSLQYGSGSLTGFFGYDTLTVQSIQVP----------
rn  TACVYHKLFDASDSSSYKHNGTELTLRYSTGTVSGFLSQDIITVGGITVT----------
cd  IACWIHHKYNSDKSSTYVKNGTSFDIHYGSGSLSGYLSQDTVSVPCQSASSASALGGVKV
ce  -ACKTHSRFQPSQSSTYSQPGQSFSIQYGTGSLSGIIGADQVSVEGLTVV----------
     **  *  .    **..         . * .*...*    * . *
     50        60        70        80        90

pp  -NQIFGLSETEPGSFLYYAPFDGILGLAYPSISASGATPVFDNLWDQGLVSQDLFSVYLS
hp  -NQIFGLSETEPGSFLYYAPFDGILGLAYPSISSSGATPVFDNIWNQGLVSQDLFSVYLS
ga  -NQEFGLSENEPGTNFVYAQFDGIMGLAYPALSVDEATTAMQGMVQEGALTSPVFSVYLS
rn  --QMFGEVTEMPALPFMLAEFDGVVGMGFIEQAIGRVTPIFDNIISQGVLKEDVFSFYYN
cd  ERQVFGEATKQPGITFIAAKFDGILGMAYPRISVNNVLPVFDNLMQQKLVDQNIFSFYLS
ce  -GQQFGESVTEPGQTFVDAEFDGILGLGYPSLAVGGVTPVFDNMMAQNLVDLPMFSVYMS
      * **     *       * ***..*. .   .             ..  .  .   .** *
     100       110       120       130       140       150

pp  SND--DSGSVVLLGGIDSSYYTGSLNWVPVSVEGYWQITLDSITMDGET-IACSGGCQAI
hp  ADD--QSGSVVIFGGIDSSYYTGSLNWVPVTVEGYWQITVDSITMNGEA-IACAEGCQAI
ga  NQQG-SSGGAVVFGGVDSSLYTGQIYWAPVTQELYWQIGIEEFLIGGQASGWCSEGCQAI
rn  RDLG----GQIVLGGSDPEHYEGNFHYINLIKTGVWQIQMKGVSVGSST-LLCEDGCLAL
cd  RDPDAQPGGELMLGGTDSKYYKGSLSYLNVTRKAYWQVHLDQVEVASGL-TLCKEGCEAI
ce  SNPEGGAGSELIFGGYDHSHFSGSLNWVPVTKQAYWQIALDNIQVGGTV-MFCSEGCQAI
              .. ** *   . *       .     **. .    .         *  ** *.
       160       170       180       190       200       210

pp  VDTGTSLLTGPTSAIANIQSDIGASENSDGEMVISCSSIDSLPDIVFTIDGVQYPLSPSA
hp  VDTGTSLLTGPTSPIANIQSDIGASENSDGDMVVSCSAISSLPDIVFTINGVQYPVPPSA
ga  VDTGTSLLTVPQQYMSALLQATGAQEDEYGQFLVNCNSIQNLPSLTFIINGVEFPLPPSS
rn  VDTGASYISGSTSSIEKLMEALGAKK-RLFDYVVKCNEGPTLPDISFHLGGKEYTLTSAD
cd  VDTGTSLMVGPVDEVRELQKAIGAVPLIQGEYMIPCEKVSTLPAITLKLGGKGYKLSPED
ce  VDTGTSLITGPSDKIKQLQNAIGAAP-VDGEYAVECANLNVMPDVTFTINGVPYTLSPTA
    ****.* .       .  .     **          . *     .* .   . *  . .
        220       230       240       250       260       270

pp  YILQDDDS----CTSGFEGMDVPTSSGE-LWILGDVFIRQYYTVFDRANNKVGLAPVA--
hp  YILQSEGS----CISGFQGMNLPTESGE-LWILGDVFIRQYFTVFDRANNQVGLAPVA--
ga  YILSNNGY----CTVGVEPTYLSSQNGQPLWILGDVFLRSYYSVYDLGNNRVGFATAA--
rn  YVFQESYSSKKLCTLAIHAMDIPPPTGP-TWALGATFIRKFYTEFDRRNNRIGFALAR--
cd  YTLKVSQAGKTLCLSGFMGMDIPPPSGP-LWILGDVFIGRYYTVFDRDNNRVGFAEAARL
ce  YTLLDFVDGMQFCSSGFQGLDIHPPAGP-LWILGDVFIRQFYSVFDRGNNRVGLAPAVP-
    *           *         .    *   * **  *.  ... .*  **..* *
        280           290       300       310       320
```

Fig. 9. Sequence alignment of human aspartic proteinases and porcine pepsin. The multiple alignment was done with the program CLUSTALV (Higgins et al., 1992). pp, porcine pepsin (Tang et al., 1973; Moravek & Kostka, 1974); hp, human pepsin (Sogawa et al., 1983); ga, gastricsin (Hayano et al., 1988); rn, renin (Imai et al., 1983); cd, cathepsin D (Faust et al., 1985); ce, cathepsin E (Azuma et al., 1989).

specificity at this position more difficult. The specificity for histidine at this position by renin is assured by the hydrogen bonding interactions made from the imidazole to Ser 76 and to Ser 222. Position 76 is occupied by glycine in all the other enzymes, and residue 222 is a threonine except in cathepsin D where it is a valine. In order for the threonine to make the equivalent hydrogen bond, it would have to adopt an unfavorable conformation. The possibility of multiple conformations may be responsible for the surprising mix of specificities observed for some cases. For example, the ability of gastricsin and cathepsin D to bind both a charged residue and a hydrophobic residue at this position indicates that these residues may be interacting with the enzyme in different ways. From the analysis of the accessible surface area and similarity in sequence, one may expect a similar P2 specificity for pepsin and cathepsin D and E. It is difficult to pinpoint the cause of the rather different profile for cathepsin D. One possibility is that in cathepsin D, the S2 region packs against a loop containing Gln 242 and Glu 244, which is disordered in pepsin. The acceptance of charged residues at P2 in gastricsin may be aided by the presence of threonine at position 289 that is either a methionine or a leucine in the other enzymes (Fig. 9).

The data for the specificity at S3 are more straightforward to interpret. The preference exhibited by pepsin for alanine over

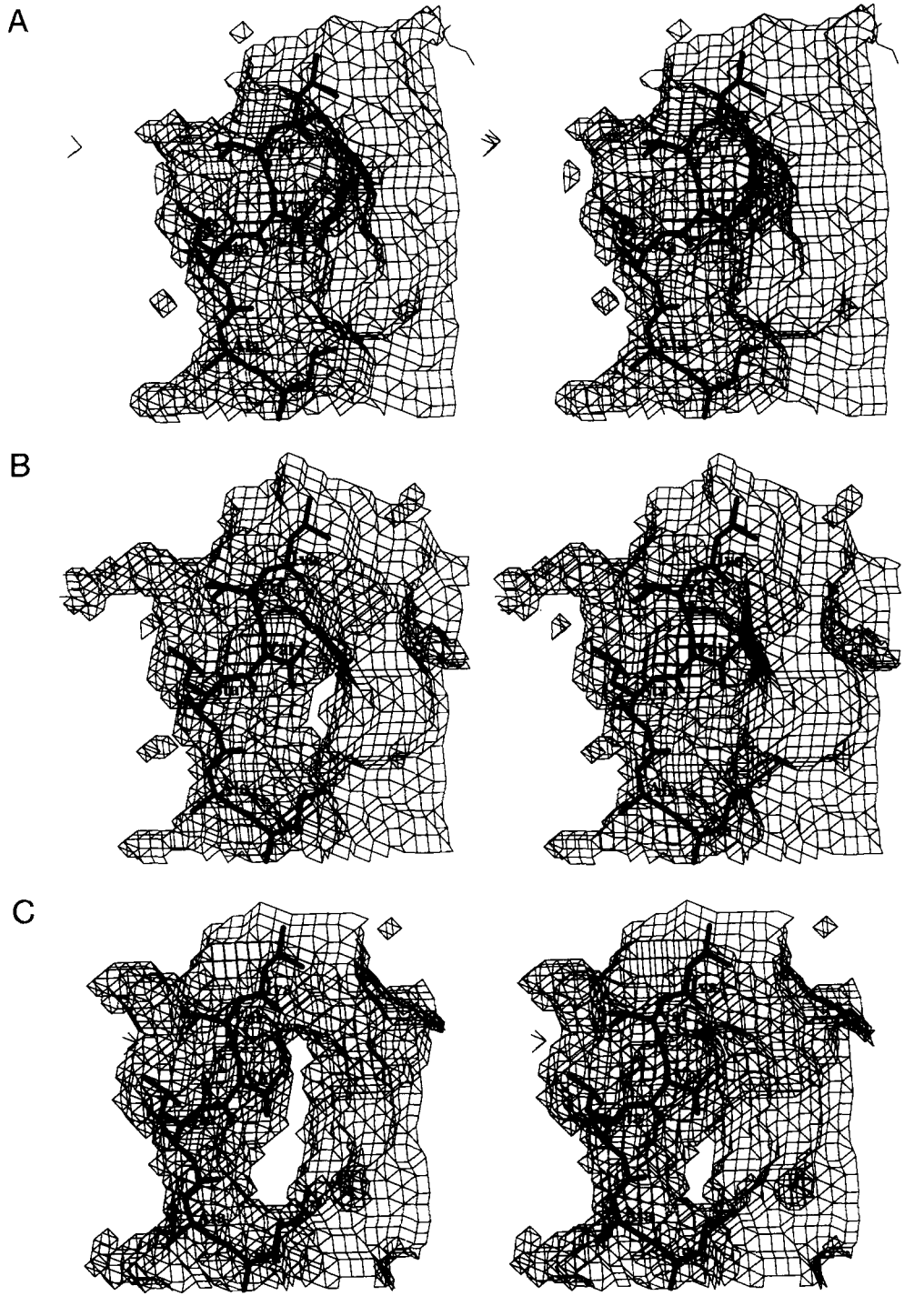

Fig. 10. Stereograms showing the solvent-accessible surfaces of (**A**) pepsin, (**B**) cathepsin D, and (**C**) renin. The molecule of pepstatin as bound to pepsin is shown as a reference in all the figures. Molecules of cathepsin D and renin were superposed on the structure of pepsin using the program O (Jones & Kjeldgaard, 1993) to facilitate the comparison.

phenylalanine is due to Phe 111, which limits severely the size of this binding site. In the other enzymes, there is a smaller residue (Asn, Thr, or Pro) at this position, allowing for the binding of a larger hydrophobic residue.

The binding sites that are C-terminal to the scissile bond are not well defined by the interactions with pepstatin due to the extra two carbon atoms inserted along the main chain of the statyl group. Nevertheless, the S1′ site should be in the neighborhood of the carbonyl oxygen atom of the central Sta; the Ala and the Sta residue that follow should occupy the S2′ and S3′ sites, respectively. As was pointed out previously, the S1′ site in renin is much smaller than in the other enzymes (Fig. 10) and this corresponds well with the presence of valine at this position in its natural substrate, angiotensinogen.

Other isoforms

Human pepsin is known to consist of up to four isoforms with differing enzymatic properties (Tarasova et al., 1994). The reported substitutions of Gln 160 to Lys, Asn 230 to Lys, and Ala 203 to Thr are not likely to modify the specificity of pepsin because these changes are located far from the binding sites. The third changes, Leu 291 to Val will likely affect the specificity at S3′ and perhaps also at S1′. All genetically evolved isoforms of pepsin can be phosphorylated at Ser 68. In either the native or the complexed pepsin structure determined here, there is no evidence of a phosphate group attached to this serine. Phosphorylation probably does not influence the activity because its location is far from the active site.

Materials and methods

Materials

Human pepsin was obtained from stomach mucosa, essentially as described previously (Tarasova et al., 1994). Stomachs were obtained during dissection no later than 3 h after death certification and kept at −70 °C before use. To start the purification, the stomachs were thawed at room temperature. Mucosa was separated with scissors; the tissue was homogenized with 4 volumes of 0.02 M Tris-HCl buffer, pH 7.2, and centrifuged at 50,000 × *g*. Pepsinogens were batch-absorbed on DEAE-cellulose DE-23 (Whatmann) and eluted in a column with 0.5 M NaCl. The zymogen peak was passed through Cibacron Blue 2GA-agarose to remove the main contaminant, serum albumin, and dialyzed against 0.02 M Tris-HCl, pH 7.2. The mixture of pepsinogens and progastricsin was obtained after chromatography on DEAE-cellulose DE-52 with elution in a gradient of 0–0.5 M NaCl. The zymogens were dialyzed against water and freeze-dried. For activation, pepsinogens were dissolved in distilled water to give a concentration of 3 mg/mL and the pH was adjusted to 3.0 with 1 M HCl. After incubation for 15 min at 37 °C, the pH was brought to 5.0 with 1 M Na acetate and the protein was dialyzed against 0.02 M piperazine-HCl, pH 5.0. Final purification of isoforms was achieved by FPLC on a 1 × 20-cm MonoQ column in a gradient of 0.15–0.3 M NaCl in 0.02 M piperazine-HCl buffer, pH 5.0. Three major isoforms were isolated, designated Pn3a, Pn3b, and Pn3c (Peek et al., 1989). The major isoform Pn3b was used for crystallization.

Pepstatin was obtained from Sigma (St. Louis, Missouri; lot 128F-0666).

Crystallization

The crystals of human pepsin A were obtained by the hanging drop technique at room temperature with $(NH_4)_2SO_4$ as the precipitant. Drops of pepsin solution (5 μL) at a protein concentration of 20 mg/mL in water were mixed with the well solution (2.5 μL) containing 40% saturated $(NH_4)_2SO_4$ in 100 mM Na acetate, pH 5.0, and 5% 2-methyl-2,4-pentanediol. Crystals appeared after approximately 1 week and grew in a few days to a maximum size of 1.2 × 0.4 × 0.2 mm.

For crystallization of the enzyme:pepstatin complex, 10 μL of pepsin solution in water at a concentration of 20 mg/mL were mixed with 2 μL of 5 mM pepstatin solution in methanol. The crystals of the pepsin:pepstatin complex grew in the same conditions as described above for the pepsin. They reached a size of 1.5 × 0.4 × 0.2 mm and were isomorphous with native crystals. The unit cells and space group are given in Table 8.

Data collection

Intensity data were collected on a San Diego multiwire detection system (Xuong et al., 1985) and the parameters of the experiment are shown in Table 8. The data were processed using the San Diego software (Howard et al., 1985). Figure 11 shows the completeness of the data collected as a function of resolution. It also shows the completeness of the data for reflections with $I > 3\sigma(I)$.

Structure solution and refinement

The structure solution was achieved by the method of molecular replacement (Rossmann, 1972), using as the search model, the structure of porcine pepsin (Sielecki et al., 1990). The fast rotation function (Crowther, 1972), as implemented in the program almn in the CCP4 suite (CCP4, 1994), was used to obtain

Table 8. *Data collection parameters*

Crystal	Native pepsin	Pepsin:pepstatin complex
Number of crystals used	1	1
Resolution of data collected	1.78 Å	1.97 Å
Total number of reflections collected	99,655	166,144
Number of unique reflections	33,057	31,184
Merging *R*	7.5%	5.4%
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell parameters	$a = 71.97$ Å $b = 151.59$ Å $c = 41.15$ Å	$a = 72.07$ Å $b = 150.97$ Å $c = 40.85$ Å

[a] X-ray source: Rigaku RU-200 rotating anode X-ray generator, 40 kV, 150 mA; graphite monochromated Cu Kα; 0.7-mm collimator.

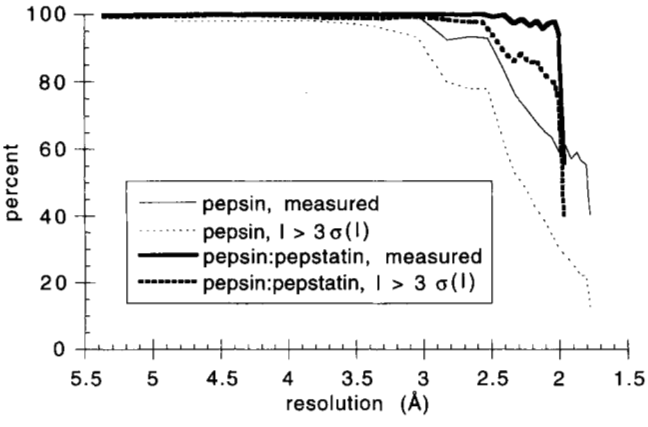

Fig. 11. Completeness of data is shown as a function of resolution. The plot also shows the percentage of reflections with $I > 3\sigma(I)$.

the rotational parameters. Data between 30 and 2.5 Å were used with the integration limits of 4–20 Å. The highest peak was obtained at 7.1σ above the mean with the second highest peak being 3.8σ above the mean. The translational parameters were obtained using the program BRUTE (Fujinaga & Read, 1987). The searches were done as two sectional searches using the screw symmetries in x and z directions. The translational search results are summarized in Table 9.

Structure refinement was done using the crystallographic extension to GROMOS (van Gunsteren & Berendsen, 1987; Fujinaga et al., 1989). Rounds of refinement interspersed with manual model refitting using FRODO (Jones, 1978) were performed. For the pepsin crystal, all data between 30 and 2.2 Å resolution were used, whereas for the pepsin:pepstatin complex, data between 30 and 2.0 Å were used.

Table 9. *Translation function results*[a]

	x	y	z	Correlation (σ)
xy section	0.167	0.158		0.413 (8.8)
	0.306	0.066		0.288 (3.3)
yz section		0.158	0.243	0.403 (8.6)
		0.158	0.122	0.285 (3.4)

[a] For each sectional search, the top two peaks are given with their coordinates in fractional units, the correlation coefficient, and the number of standard deviations above the mean.

Acknowledgments

This research was sponsored in part by the Medical Research Council of Canada and in part by the National Cancer Institute, DHHS, under contract NO1-CO-46000 with ABL.

References

Abad-Zapatero C, Rydel TJ, Erickson J. 1990. Revised 2.3 Å structure of porcine pepsin: Evidence for a flexible subdomain. *Proteins Struct Funct Genet 8*:62–81.

Azuma T, Pals G, Mohandas TK, Couvreur JM, Taggart RT. 1989. Human gastric cathepsin E. Predicted sequence localization to chromosome 1, and sequence homology with other aspartic proteinases. *J Biol Chem 264*:16748–16753.

Bailey D, Cooper JB, Veerapandian B, Blundell TL, Atrash B, Jones DM, Szelke M. 1993. X-ray-crystallographic studies of complexes of pepstatin A and a statine-containing human renin inhibitor with endothiapepsin. *Biochem J 289*:363–371.

Baldwin ET, Bhat TN, Gulnik S, Hosur MV, Sowder RC II, Cachau RE, Collins J, Silva AM, Erickson JW. 1993. Crystal structures of native and inhibited forms of human cathepsin D: Implications for lysosomal targeting and drug design. *Proc Natl Acad Sci USA 90*:6796–6800.

Baxter A, Campbell CJ, Grinham CJ, Keane RM, Lawton BC, Pendlebury JE. 1990. Substrate and inhibitor studies with human gastric aspartic proteinases. *Biochem J 267*:665–669.

Bennett K, Levine T, Ellis JS, Peanasky RJ, Samloff IM, Kay J, Chain BM. 1992. Antigen processing for presentation by class II major histocompatibility complex requires cleavage by cathepsin E. *Eur J Immunol 22*: 1519–1524.

Blundell TL, Jenkins J, Pearl LH, Sewell T. 1985. The high resolution structure of endothiapepsin. In: Kostka V, ed. *Aspartic proteinases and their inhibitors*. Berlin: Walter de Gruyter. pp 151–161.

Bott R, Subramanian E, Davies D. 1982. Three-dimensional structure of the complex of the *Rhizopus chinensis* carboxyl proteinase and pepstatin at 2.5 Å resolution. *Biochemistry 21*:6956–6962.

Cataldo AM, Nixon RA. 1990. Enzymatically active lysosomal proteases are associated with amyloid deposits in Alzheimer brain. *Proc Natl Acad Sci USA 87*:3861–3865.

CCP4. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D 50*:760–763.

Cody RJ. 1994. The clinical potential of renin inhibitors and angiotensin antagonists. *Drugs 47*:586–598.

Crowther RA. 1972. The fast rotation function. In: Rossmann MG, ed. *The molecular replacement method*. New York: Gordon and Breach. pp 173–178.

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

Epps DE, Cheney J, Schostarez H, Sawyer TK, Prairie M, Krueger WC, Mandel F. 1990. Thermodynamics of the interaction of inhibitors with the binding site of recombinant human renin. *J Med Chem 33*:2080–2086.

Evers MPJ, Zelle B, Bebelman JP, van Beusechem V, Kraakman L, Hoffer MJV, Pronk JC, Mager WH, Planta RJ, Eriksson AW, Frants RR. 1989. Nucleotide sequence comparison of five human pepsinogen A (PGA) genes: Evolution of the PGA multigene family. *Genomics 4*:232–239.

Evers MPJ, Zelle B, Peeper DS, Mager WH, Planta RJ, Eriksson AW, Frants RR. 1988. Genomic structure and evolution of the human pepsinogen A multigene family. *Hum Genet 78*:79–82.

Faust PL, Kornfeld S, Chirgwin JM. 1985. Cloning and sequence analysis of cDNA for human cathepsin D. *Proc Natl Acad Sci USA 82*:4910–4914.

Foltmann B. 1981. Gastric proteinases—Structure, function, evolution and mechanism of action. *Essays Biochem 17*:52–84.

Foundling SI, Cooper J, Watson FE, Cleasby A, Pearl LH, Sibanda BL, Hemmings A, Wood SP, Blundell TL, Valler MJ, Norey CG, Kay J, Boger J, Dunn BM, Leckie BJ, Jones DM, Atrash B, Hallet A, Szelke M. 1987. High resolution X-ray analysis of renin inhibitor-aspartic proteinase complexes. *Nature 327*:349–352.

Fruton JS. 1976. The mechanism of the catalytic action of pepsin and related acid proteinases. *Adv Enzymol 44*:1–36.

Fujinaga M, Gros P, van Gunsteren WF. 1989. Testing the method of crystallographic refinement using molecular dynamics. *J Appl Crystallogr 22*:1–8.

Fujinaga M, Read RJ. 1987. Experiences with a new translation-function program. *J Appl Crystallogr 20*:517–521.

Hayano T, Sogawa K, Ichihara Y, Fujii-Kuriyama Y, Takahashi K. 1988. Primary structure of human pepsinogen C gene. *J Biol Chem 263*:1382–1385.

Higgins DG, Bleasby AJ, Fuchs R. 1992. CLUSTAL V: Improved software for multiple sequence alignment. *Comput Applic Biosci 8*:189–191.

Howard AJ, Nielsen C, Xuong NH. 1985. Software for a diffractometer with multiwire area detector. *Methods Enzymol 114*:452–472.

Imai T, Miyazaki H, Hirose S, Hori H, Hayashi T, Kageyama R, Ohkubo H, Nakanishi S, Murakami K. 1983. Cloning and sequence analysis of cDNA for human renin precursor. *Proc Natl Acad Sci USA 80*:7405–7409.

James MNG, Sielecki AR. 1983. Structure and refinement of penicillopepsin at 1.8 Å resolution. *J Mol Biol 163*:299–361.

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

Jones TA, Kjeldgaard M. 1993. *O. 5.9*. Uppsala, Sweden: DatONO AB.

Jones TA, 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.

Jupp RA, Dunn BM, Jacobs JW, Vlasuk G, Arcuri KE, Veber DF, Perlow DS, Payne LS, Boger J, De Laszlo S, Chakravarty PK, Ten Broeke J, Hangauer DG, Ondeyka D, Greenlee WJ, Kay J. 1990. The selectivity of statine-based inhibitors against various human aspartic proteinases. *Biochem J 265*:871–878.

Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*:2577–2637.

Kleywegt GJ, Jones TA. 1994. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr D 50*:178–185.

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.

Lees WE, Kalinka S, Meech JD, Capper SJ, Cook ND, Kay J. 1990. Generation of human endothelin by cathepsin E. *FEBS Lett 273*:99–102.

Moravek L, Kostka V. 1974. Complete amino acid sequence of hog pepsin. *FEBS Lett 43*:207–211.

Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. 1992. Stereochemical quality of protein structure coordinates. *Proteins Struct Funct Genet 12*:345–364.

Peek K, Roberts NB, Taylor WH. 1989. Improved separation of human pepsins from gastric juice by high-performance ion-exchange chromatography. *J Chromatogr 476*:291–297.

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

Rochefort H. 1990. Biological and clinical significance of cathepsin D in breast cancer. *Semin Cancer Biol 1*:153–160.

Rossmann MG, ed. 1972. *The molecular replacement method*. New York: Gordon and Breach.

Šali A, Veerapandian B, Cooper JB, Foundling SI, Hoover DJ, Blundell TL. 1989. High-resolution X-ray diffraction study of the complex between endothiapepsin and an oligopeptide inhibitor: The analysis of the inhibitor binding and description of the rigid body shift in the enzyme. *EMBO J 8*:2179–2188.

Šali A, Veerapandian B, Cooper JB, Moss DS, Hofmann T, Blundell TL. 1992. Domain flexibility in aspartic proteinases. *Proteins Struct Funct Genet 12*:158–170.

Samloff IM. 1969. Slow moving protease and the seven pepsinogens. Electrophoretic demonstration of the existence of eight proteolytic fractions in human gastric mucosa. *Gastroenterology 57*:659–669.

Scarborough PE, Guruprasad K, Topham C, Richo GR, Conner GE, Blundell TL, Dunn BM. 1993. Exploration of subsite binding specificity of human cathepsin D through kinetics and rule-based molecular modeling. *Protein Sci 2*:264–276.

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

Sielecki AR, Fedorov AA, Boodhoo A, Andreeva NS, James MNG. 1990. Molecular and crystal structures of monoclinic porcine pepsin refined at 1.8 Å resolution. *J Mol Biol 214*:143–170.

Sogawa K, Fujii-Kuriyama Y, Mizukami Y, Ichihara Y, Takahashi K. 1983. Primary structure of human pepsinogen gene. *J Biol Chem 258*:5306–5311.

Suguna K, Padlan EA, Bott R, Boger J, Parris KD, Davies DR. 1992. Structures of complexes of rhizopuspepsin with pepstatin and other statine-containing inhibitors. *Proteins Struct Funct Genet 13*:195–205.

Tang J, Sepulveda P, Marciniszyn JJ, Chen KC, Huang WY, Tao N, Liu D, Lanier JP. 1973. Amino-acid sequence of porcine pepsin. *Proc Natl Acad Sci USA 70*:3437–3439.

Tarasova N, Denslow ND, Parten BF, Tran N, Nhuyen HP, Jones A, Roberts NB, Dunn BM. 1994. A comparative study of amino acid sequences of three major isoforms of human pepsin A. In: Takahashi K, ed. *Aspartic proteases and their inhibitors*. New York: Plenum Press. Forthcoming.

Tarasova N, Foltmann B, Szecsi P. 1986. An aspartic protease from human erythrocyte is immunologically indistinguishable from a non-pepsin, electrophoretically slow moving protease from gastric mucosa. *Biochim Biophys Acta 869*:96–100.

van Gunsteren WF, Berendsen HJC. 1987. *GROMOS. Groningen molecular simulation library*. Groningen, The Netherlands: BIOMOS b.v.

Xuong NH, Sullivan D, Neilsen C, Hamlin R. 1985. Use of the multiwire area detector diffractometer as a national resource for protein crystallography. *Acta Crystallogr B 41*:367–269.