PROTEINS: Structure, Function, and Genetics 8:62–81 (1990)

Revised 2.3 Å Structure of Porcine Pepsin: Evidence for a Flexible Subdomain

Cele Abad-Zapatero, Timothy J. Rydel, and John Erickson
Protein Crystallography Laboratory, D-47E, Abbott Laboratories, Abbott Park, Illinois 60064

ABSTRACT **A revised three-dimensional crystal structure of ethanol-inhibited porcine pepsin refined to an *R*-factor of 0.171 at 2.3 Å resolution is presented and compared to the refined structures of the fungal aspartic proteinases: penicillopepsin, rhizopuspepsin, and endothiapepsin. Pepsin is composed of two nearly equal N and C domains related by an intra dyad. The overall polypeptide fold and active site structures are homologous for pepsin and the fungal enzymes. The weak inhibition of pepsin by ethanol can be explained by the presence of one or more ethanol molecules, in the vicinity of the active site carboxylates, which slightly alter the hydrogen-bonding network and which may compete with substrate binding in the active site. Structural superposition analysis showed that the N domains aligned better than the C-domains for pepsin and the fungal aspartic proteinases: 107–140 C_α pairs aligned to 0.72–0.85 Å rms for the N domains; 64–95 C_α pairs aligned to 0.78–1.03 Å rms for the C domains. The major structural difference between pepsin and the fungal enzymes concerns a newly described subdomain whose conformation varies markedly among these enzyme structures. The subdomain in pepsin comprises nearly 100 residues and is composed of two contiguous segments within the C domain (residues 192–212 and 223–299). the subdomain is connected, or "hinged," to a mixed β-sheet that forms one of the structurally invariant, active site ψ-loops. Relative subdomain displacements as large as a 21.0° rotation and a 5.9 Å translation were observed among the different enzymes. There is some suggestion in pepsin that the subdomain may be flexible and perhaps plays a structural role in mediating substrate binding, determining the substrate specificity, or in the activation of the zymogen.**

Key words: pepsin, aspartic proteinases, subdomains, structure comparison

INTRODUCTION

The aspartic proteinases comprise an important family of proteolytic enzymes that are distinguished by the presence of two conserved aspartic acid residues in the active site.[1,2] Aspartic proteinases are widespread in nature and several have long been known to be medically (renin, cathepsins D and E, pepsin) or commercially (chymosin) important. Current developments in the structure and function of these enzymes as well as their inhibitors have been reviewed.[3] Recently, it has been shown that the proteases encoded by retroviruses are structurally related to the eukaryotic aspartic proteinases and thus represent novel therapeutic targets for the design of antiretroviral drugs.[4–7]

Central to our knowledge of aspartic proteinase function is a detailed structural description of these enzymes. The X-ray crystal structures of several aspartic proteinases from fungi have been refined: penicillopepsin[8] at 1.8 Å, endothia parasitica pepsin[9] at 2.1 Å, and rhizopus chinensis protease[10] at 1.8 Å. The 2.5 Å crystal structure of recombinant human renin was reported recently, but coordinates for this structure are not currently available.[11] A 3.0 Å structure of porcine pepsin was described perviously,[12] and a partially refined structure at 2.0 Å resolution has been reported more recently.[13] However, both of the available pepsin structures display suboptimal stereochemistry, and are limited in their usefulness. We undertook to recollect diffraction data and to improve the pepsin crystal structure for several reasons. First, we wished to provide a suitable pepsin structure which may be used in molecular replacement techniques to help solve the structures of nonisomorphous crystals of pepsin/renin inhibitor complexes which we have obtained.[14,15] Second, pepsin exhibits nearly twice the sequence homology to renin as do the fungal proteases, and thus pepsin should be a more suitable structure upon which to construct renin models.[16,17] Finally, a well-refined pepsin structure should provide a structural framework for interpreting studies concerned with the mechanisms of pepsin catalysis and zymogen activation, and for evolutionary analysis of aspartic proteinase structure.

Received December 20, 1989; revision accepted February 28, 1990.

Address reprint requests to John Erickson, Protein Crystallography Laboratory, D-47E, Abbott Laboratories, 1 Abbott Park Road, Abbott Park, IL 60064.
Timothy J. Rydel's present address: Department of Chemistry, Michigan State University, East Lansing, MI 48824.

© 1990 WILEY-LISS, INC.

We have refined the crystal structure of porcine pepsin at 2.3 Å resolution to an R-value of 0.171 based on new crystallographic data. We present here a description of the revised pepsin structure, and its comparison with the fungal aspartic proteinases.

EXPERIMENTAL PROCEDURES

Crystallization

Porcine pepsin was obtained from Sigma Chemical Co. The enzyme was repurified by ion-exchange chromatography and stored at −20°C as a lyophilized powder. Pepsin was crystallized from a 20% ethanol/water solution adjusted to pH 2.0 as described.[13] The monoclinic $P2_1$ crystals (a = 55.2 Å, b = 36.4 Å, c = 73.6 Å, β = 104.0°) grew as thin plates in rosettes and were very difficult to handle. The cell constants varied between crystals and occasionally varied during data collection from a single crystal by as much as 4% along the c axis. For these reasons, only data from crystals which remained stable during the data collection were used for the refinement studies reported here.

Data Collection

Diffraction data to 3.0 Å resolution were collected from a single crystal (a = 55.2 Å, b = 36.4 Å, c = 73.6 Å, β = 104.0°) by diffractometry using a Rigaku AFC5 diffractometer equipped with a graphite crystal monochromator. CuK_α X-rays were generated by a Rigaku RU-200 rotating anode operated at 50 kV and 110 mA. Data were collected using an ω-scan in which background was measured on both sides of the diffraction peak for equal intervals, the total of which corresponded to about 50% of the total peak scan time. Background-corrected intensities were reduced to $|F|$'s after applying Lorentz and polarization corrections. Decay was monitored by measuring a set of nine reflections throughout the course of data collection at regularly selected intervals, usually after a block of 150–200 reflections. Decay was computed as a function of both time and resolution by fitting the standard $|F|$'s to an exponential decay curve. Further data processing and Fourier calculations were done using the program PROTEIN. A total of 6018 unique reflections were collected from ∞ – 3.0 Å with an R_{sym} = 0.032 for 432 symmetry-related reflections. A total of 3225 additional observations between 3.1 and 2.5 Å resolution were collected from a second pepsin crystal and were merged with the 3.0 Å data. The final data set contained 9502 observations which reduced to 8557 unique observations with an R-merge of 0.082 for the 1659 redundant reflections. This data set was essentially complete to 3.0 Å (> 90% observed) and contained over 70% of the possible reflections between 3.0 and 2.8 Å.

A 2.0 Å resolution data set was collected from a single crystal (a = 55.3 Å, b = 36.4 Å, c = 73.8 Å, β = 103.4°) using a Xentronics multiwire area detector at the Argonne Midwest Area Detector Center. The indexing and integration of the reflection intensities were performed using the XENGEN software package.[18] A total scan of 180° in ω, at a χ setting of 30°, was performed in three 60° ω sweeps with a 60° shift in φ between sweeps. Crystal-to-detector distance was 10 cm at a swing angle (2θ) of 19°. Exposure per frame was 200 seconds for each 0.25° scan. A total of 24,237 reflections were measured resulting in 17,211 unique reflections to 1.97 Å resolution (83% complete) with an R_{merge} of 0.064. The area detector data were strong up to 2.5 Å ($\langle I/\sigma(I)\rangle \geq$ 4.5) but were considerably weaker between 2.5 and 2.20 Å ($\langle I/\sigma(I)\rangle = 2.2$) and between 2.21 and 1.97 Å ($\langle I/\sigma(I)\rangle = 1.2$). A total of 12,252 unique reflections with an $F/\sigma(F) \geq 2$ were selected from the area detector data set. These data represented 62% of the theoretically possible observations to 2.0 Å. The diffractometer data scaled poorly ($R_{merge} = 0.145$) with the area detector data and therefore only the latter was used for refinement at high resolution. Since the observed reflections beyond 2.3 Å were very weak and were observed only in a limited number (<50% of the theoretically possible), the final stages of the refinement used only 8,742 reflections between 5 and 2.3 Å resolution (77% complete).

Refinement

Refinement of the initial pepsin structure proceeded in four stages. First, the initial pepsin model[12] (R = 0.45) was regularized using CORELS[19] and subsequently revised using FRODO[20] against $2F_o - F_c$ and $3F_o - 2F_c$ electron density maps computed using the 10.0–3.0 Å diffractometer data. Two rounds of manual revision and refinement using CORELS and PROLSQ[21] with no solvent molecules reduced the R-factor to 0.297 for data 6.0 and 3.0 Å. In the second stage, diffractometer data to 2.6 Å were added and, after a round of revision/refinement, the R-factor dropped to 0.278 (5.0–2.6 ; no solvent). In the third stage, the structure was refined using the area detector data. From this point onward, refinement was performed by simulated annealing and energy minimization using the program XPLOR and a reannealing strategy to be described more fully elsewhere.[22] After two rounds of XPLOR with one intervening round of manual revision and the addition of over 200 water molecules, the R-factor was reduced to 0.153 (5.0–2.5 Å) with reasonable stereochemistry (rms bonds = 0.020 Å and rms angles = 4.275°). In the fourth stage, area detector data from 2.5 to 2.3 Å ($F/\sigma(F) > 2.0$) were included after manual scaling of F_{obs} to F_{calc}. The additional data allowed definite improvements to be made to the structure judging from the final stereochemistry (Table I and Fig. 1). After several rounds of XPLOR, followed by PROLSQ, the R-factor was 0.171 for 8742 reflections from 5.0 to 2.3 Å. A plot of the R-factor as a function of resolu-

TABLE I. Summary of Stereochemical Properties for the 2.3 Å Structure of Porcine Pepsin

	Target value	Final value
R-factor		0.171
Resolution range (Å)		5.0–2.3
rms coordinate shift (Å)		0.015
rms B shift (Å^2)		−0.013
Average $\|F_{obs}\| - \|F_{calc}\|$	3.95	3.80
Distances (Å)*		
$i - (i + 1)$	0.020	0.018
$i - (i + 2)$	0.040	0.052
$i - (i + 3)$	0.050	0.064
Planarity (Å)	0.020	0.018
Chiral volume (Å^3)	0.150	0.165
Torsion angles (deg)		
Planar	3.0	5.5
Staggered	15.0	19.2
Orthonormal	20.0	19.5
Possible H-bonds (Å)	0.5	0.101
Temperature factors (Å^2)		
Main chain 1–2	1.0	2.4
Main chain 1–3	1.5	3.9
Side chain 1–2	1.5	2.8
Side chain 1–3	2.0	4.2
No. of reflections		8742
Protein atoms		2429
H_2O molecules		206

*$i - (i + 1)$, bond distance; $i - (i + 2)$, next nearest neighbor distance of three bonded atoms defining a bond angle; $i - (i + 3)$, the planar 1–4 distance.

tion is shown in Figure 2. A complete coordinate set of the final pepsin structure (3PEP) has been deposited with the Protein Data Bank.[23]

Structural Comparisons

Atomic coordinates for rhizopuspepsin (2APR),[10] penicillopepsin (2APP),[8] and endothiapepsin (4APE)[9] are available from the Protein Data Bank.[23] A set of coordinates for endothiapepsin which had been refined further to 1.8 Å was kindly supplied by Tom Blundell. Structural alignments were performed using the superposition method of Rossmann and Argos[24] and the following input parameters: E1 = E2 = 5.00; PCUT = 0.1; ESIG = 1.0; and TSIG = 0.5. In all cases, convergence occurred within less than 10 cycles. Final E1 and E2 values ranged from 0.84 to 1.02 and 1.04 to 1.27, respectively, for the pepsin vs. fungal proteinase superpositions. For the internal symmetry analysis of the pepsin structure, E1 and E2 values ranged from 2.6 to 2.8 and 3.3 to 3.4, respectively.

RESULTS AND DISCUSSION

The Revised Structure

A schematic diagram of the bilobal, croissant-shaped pepsin molecule is shown in Figure 3 and a stereo diagram of the C_α atom tracing is presented in Figure 4. These views illustrate the overall secondary structure and chain folding of pepsin with its two main domains: the N domain consisting of residues 1–176, and the C domain formed by residues 177–326. Each domain in pepsin contains one of the two catalytically important aspartic acid residues (Asp-32 and Asp-215). The overall temperature factor, B, for the pepsin structure is 25 Å^2. The mean temperature factor of the amino acid residues for the segments 170–205 and 230–295 was significantly larger ($B > 35$ Å^2). All regions of the polypeptide chain were well-defined in the $2F_o$-F_c electron density map except for the β turn comprising Asp-278–Ser-281 and the chain segment corresponding to residues Asp-290–Glu-297. These two regions exhibited the highest isotropic temperature factors ($B > 60$ Å^2) and are probably disordered in the crystal structure. The latter region represents a highly variable sequence among the aspartic proteinases, and it contains a 4 or 5 residue insertion in pepsin (and human renin) relative to the fungal enzymes as well as to chymosin.

The revised model consists of 326 amino acid residues, one less than for the previously described structure.[12,13] The deletion occurs at Ile-230, in the middle of the longest α-helical segment (residues 225–235). This interpretation agrees with the amino acid sequence of Moravek and Kostka[25] as well as that deduced from the nucleotide sequence of the porcine pepsinogen gene.[26] We observed a *cis* peptide bond between Thr-22 and Pro-23, in a five residue β turn, immediately following the βB strand. A *cis* peptide bond is also present in an analogous position in rhizopuspepsin[10] (between Thr-25 and Pro-26). The present model does not support the Asp to Tyr substitution at position 242 which was recently reported for the porcine pepsinogen gene sequence by Lin and co-workers.[26] The current model has preliminary positions for 206 water molecules of which approximately 20 are located in the active site cleft. Around 10% of the electron density maxima associated with the putative water oxygen atoms are larger and more elongated than expected for water molecules and are probably ethanol molecules.

While the overall structure and topology of the revised model of porcine pepsin generally agree with those for the previously reported structure, several substantial differences were found (Fig. 4b). Residues 2–9 in the N-terminus of pepsin form a regular β-strand (βA1 and βA2; Fig. 3) as part of the intradomain sheet. In the earlier model, this region adopted an irregular conformation, with the N-terminus folded back on itself to form an ion-pair with Glu-4. Several major structural differences were found throughout the C-domain of the molecule, particularly in the region comprising residues Ile-231–Glu-297. These differences consisted of substantially altered loop conformations, and the introduction of new secondary structure elements. Besides these localized structural differences, this entire segment of the molecule is apparently shifted relative to the

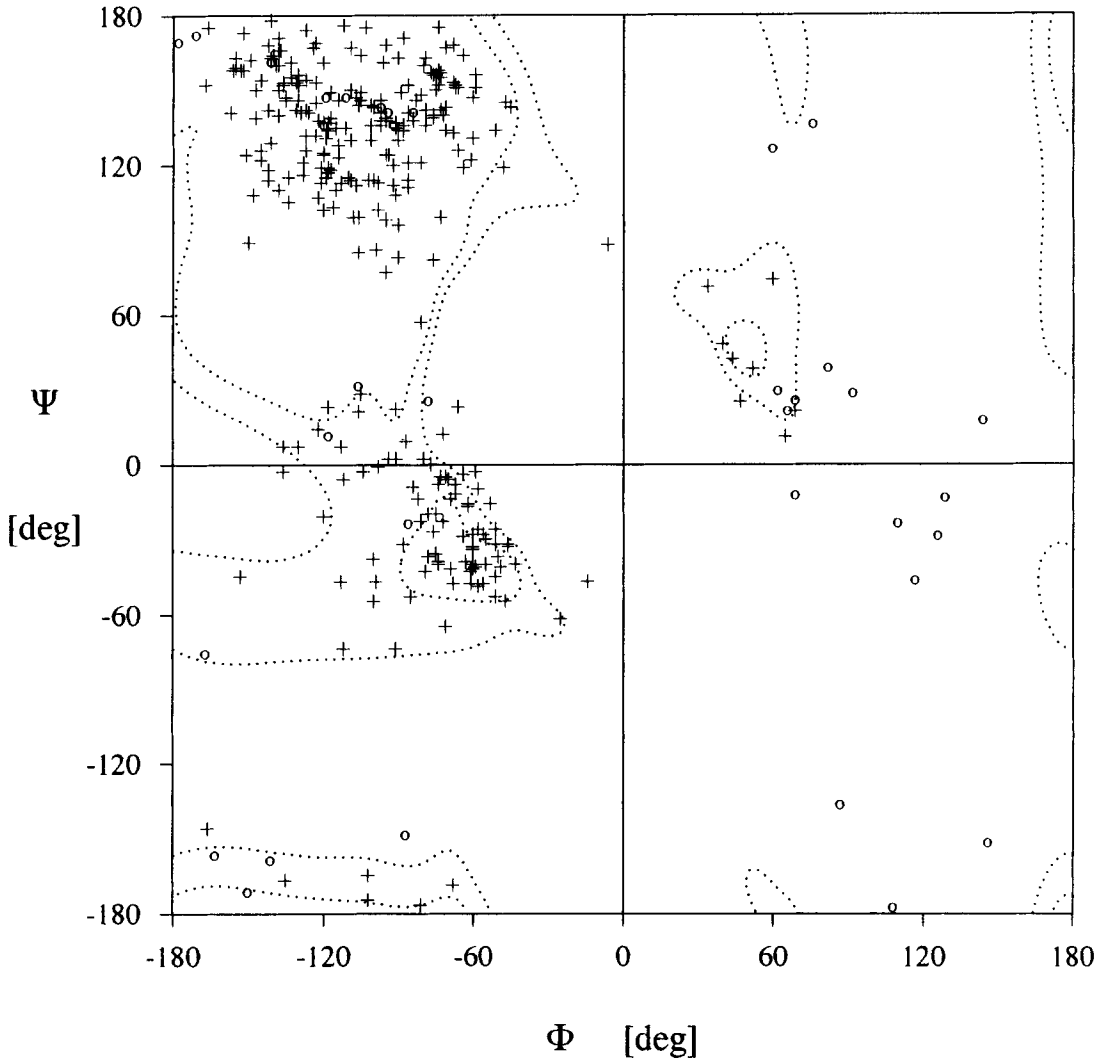

Fig. 1. Ramachandran (ϕ,ψ) plots for the 326 amino acid residues of pepsin. Crosses and open circles indicate conformational angles for the 291 nonglycine and 35 glycine residues, respectively. The isoenergy contours correspond to 4 and 8 kcal/mol based on calculations for the alanine dipeptide.[40]

previous pepsin model. Overall, the revised pepsin structure differs from the previous model by 3.0 Å rms for 326 C_a atom pairs.

Secondary Structure

All aspartic proteinases known to date belong to the class of "all β" proteins.[28] Their β-sheet topology has been described in detail for the fungal enzymes[8–10] and compared to the initial pepsin models by Andreeva and co-workers.[12,13,27] The structural elements and hydrogen bonding pattern for pepsin are illustrated in Figure 5 (see also Figs. 3 and 4a). Approximately half of the residues adopt conformational angles (ϕ,ψ) corresponding to parallel or antiparallel β-sheets (Fig. 1). Overall, the structure consists of one interdomain sheet (I) and six intradomain sheets (II–VII) (Table II). The interdomain sheet is the most regular and well-defined, and comprises six antiparallel strands. Sheets II and III are related by the interdomain dyad which intersects the plane of sheet I. One of the strands of sheet III is interrupted by the disulfide bond between Cys-206 and Cys-210 which is not present in the fungal enzymes. Sheet IV, which has no symmetry-related counterpart, contains the active site "flap," a β-hairpin turn consisting of residues Thr-74–Gly-78. Sheet V is formed by two nearly perpendicular pairs of antiparallel β-strands which share only two main chain hydrogen bonds (Met-245 O–N Ser-284 and Ile-247 N–O Cys-282), and which are interconnected through a conserved disulfide bond (Cys-249–Cys-282). This disulfide bridge may serve to stabilize the unusually high twist of this sheet. In addition to the antiparallel β-sheets I–V, we also observed a rather

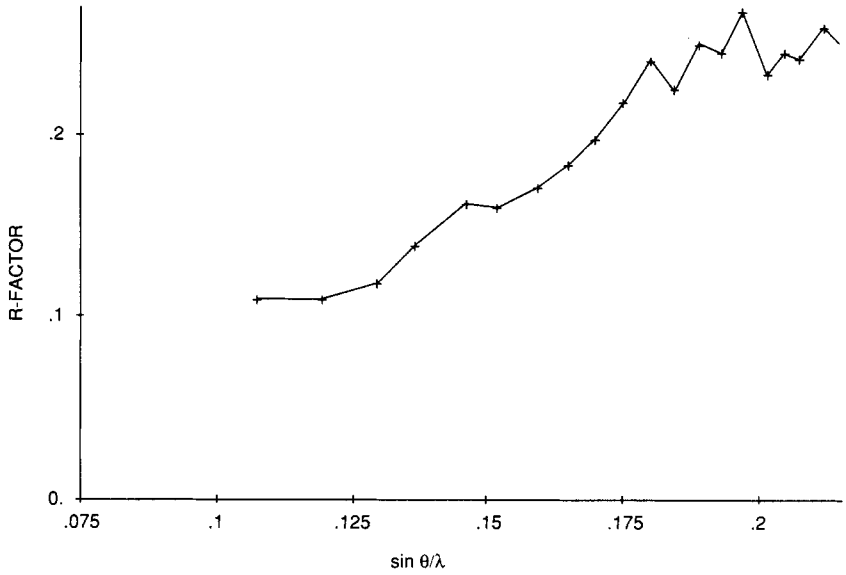

Fig. 2. R-factor as a function of sin θ/λ.

extensive, five-stranded mixed β-sheet structure (VI) in the C domain. Sheet VI contains the active site ψ_2 sheet flanked on either side by one additional β-strand. In contrast, the corresponding β-sheet in rhizopuspepsin contains only four strands. Sheet VII, a six-stranded, mixed β-sheet, contains the active site ψ_1 sheet in addition to three other short strands (βE_2, βF_2, βG_2) which extend toward the periphery of the molecule.

Pepsin contains seven α-helical segments of varying lengths (Table III). The four shortest ones occur as two pairs (αA and αB; αE and αF), one pair in each domain of the polypeptide chain. Only one member of each pair (αB and αF) is present in endothiapepsin and penicillopepsin, but three of the four helices (αB, αE, and αF) are conserved in rhizopuspepsin. The longest helix, αD (residues 225–235), is exposed to solvent and connects sheets V and VI. It is preserved among the fungal proteinases but its orientation and position relative to the rest of the structure vary. Helices αC and αG, which are related by an interdomain 2-fold (see below) are well conserved among the fungal proteinases, particularly αG which is buried and forms part of the C-domain ψ_2-loop.

Polar Side Chain Interactions

A total of 28 side chain interactions between polar residues were observed in the refined pepsin structure (Table IV). While there are 47 ionizable side chains in pepsin (43 acidic and 4 basic residues), only 4 of the 28 contacts were between ionizable groups. Of these 4 contacts, 2 were between pairs of carboxylates (Asp-54–Asp-118 and Asp-290–Glu-297)and presumably represent hydrogen-bonded interactions. These interactions are presumably enhanced in the low pH crystal structure where one or both carboxylates in each pair of contacts may be protonated. There are two interdomain salt bridges in 2-fold related pepsin—one between Asp-11 and Arg-307, and the other between Asp-138 and Arg-315. The first interdomain bridge is essentially conserved in all the fungal enzymes but with Arg-307 being replaced by a lysine residue in all three structures. On the other hand, only the acidic member of the pair is conserved in the second salt bridge. The amino acid residue at the position corresponding to Arg-315 in the fungal enzymes is glycine in endothiapepsin, serine in penicillopepsin, and glutamine in rhizopuspepsin. Therefore, this second salt bridge seems to be a novel interaction in pepsin. The relative importance of these ion pairs for the structure and function of the enzyme has still to be ascertained.

In the current model, the carboxylate side chain of Glu-4 is hydrogen-bonded to the carbonyl oxygen of Thr-17 and cannot interact with the amino-terminus of the polypeptide as had been proposed for the previous pepsin model.[13] Instead, the amino group of residue Ile-1 appears to share a solvent molecule with the carbonyl oxygen of Leu-145 (data not shown).

Internal Symmetry

Structural analyses of the aspartic proteinases has led to the proposal that these enzymes may have evolved by gene duplication from an ancestral protein of about 150 residues whose fold constitutes one of the pepsin "lobes." Further analysis suggested that each lobe may itself be a product of gene

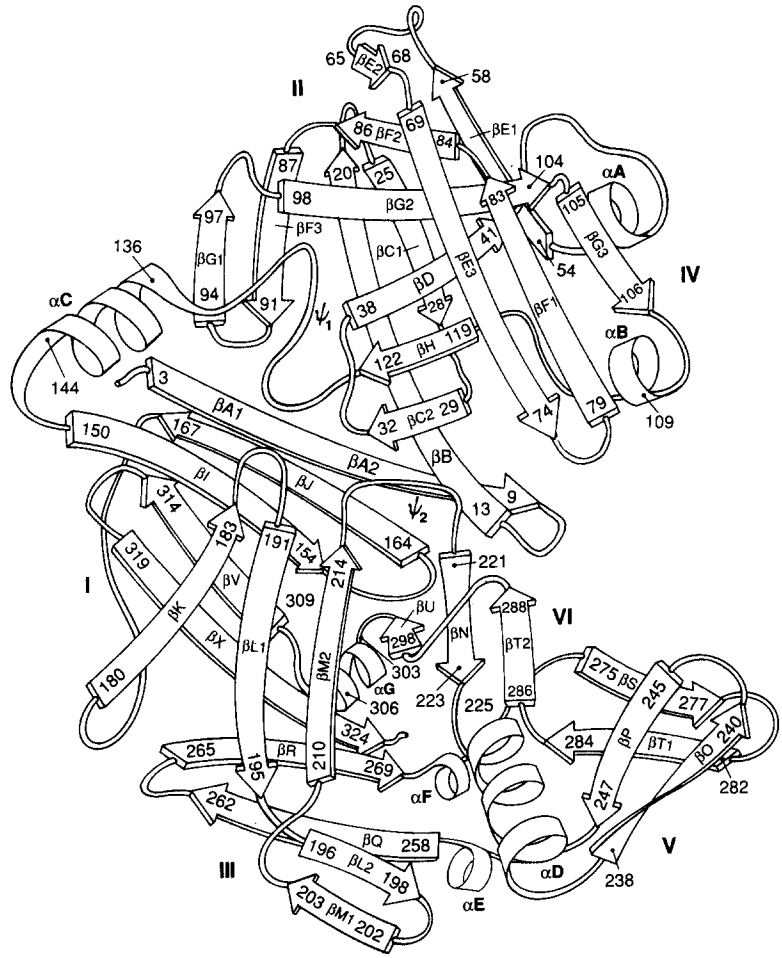

Fig. 3. Schematic diagram of the polypeptide fold of porcine pepsin. The approximate boundaries for elements of secondary structure are labeled along the chain. The most prominent β-sheets are labeled using Roman numerals following the convention used for penicillopepsin[8] and rhizopuspepsin.[10] (Diagram adapted from James and Sielecki.[8])

duplication.[27,29] Hence, both interdomain and intradomain dyad symmetry can be detected in these enzyme structures. Superposition of the N and C domains of pepsin resulted in 70 pairs of C_α atoms which are related by a rotation of 174.3° with an rms deviation of 2.4 Å. The regions of topological equivalence are similar to those found in the fungal proteinases and are illustrated in Figures 6 and 7. The sequence identity for all the equivalenced residues is 25%. The equivalenced portions of the molecule with the most extensive sequence homology (10/22 residues conserved) comprise the active site ψ_1 and ψ_2 sheets[10] in the proximity of the interdomain dyad: Asp-26–Ser-42 equivalent to Gly-209–Thr-225; Asp-118–Gly-122 equivalent to Leu-297–Gly-302. These 22 C_α pairs agree to an overall rms of 1.6 Å and are related by a rotation of 178.4°. A comparison of the interdomain symmetry of porcine pepsin with that of the fungal proteinases is listed in Table V.

The intradomain symmetry for pepsin is much less extensive than the interdomain symmetry, and differs slightly between the N and C domains. In the N domain, an approximate 2-fold axis (175.4° rotation) superimposes a total of 26 C_α atom pairs from four segments with an overall rms deviation of 2.4 Å. The equivalent C_αs are as follows: Tyr-14–Ile-20 and Ile-83–Val-89; Gly-21–Gln-25 and Gly-92–Asp-96; Asp-26–Gly-34 and Asn-98–Thr-106; and Leu-38–Ser-42 and Gly-119–Leu-123. The approximate dyad (170.0° rotation) for the C domain comprises only 23 equivalent C_α pairs to an rms deviation of 2.7 Å: Asp-195–Ile-204 and Ile-258–Tyr-267; Gln-211–Thr-218 and Gly-285–Pro-292; and Leu-220–Pro-224 and Leu-298–Gly-302. These residues largely comprise those same residues that are also related by the interdomain dyad, thus supporting the view that eukaryotic aspartic proteinases have evolved by one or more duplication events of a primordial gene. The differences between the N- and C-intradomain symmetries probably reflect different constraints on the evolution of these two domains. The gene duplication hypothesis has been

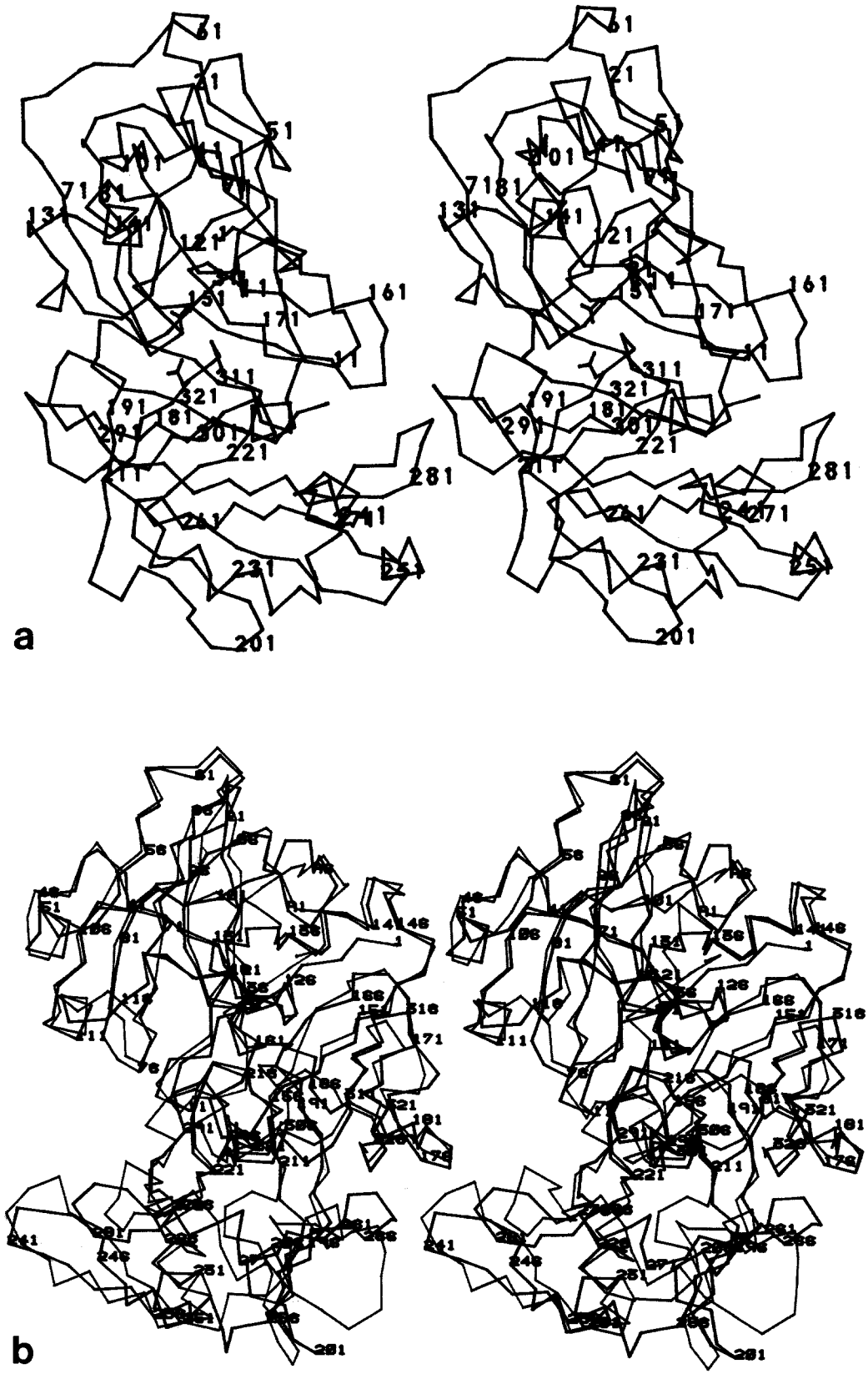

Fig. 4. Stereo diagrams of the revised structure of porcine pepsin. (**a**) C_α atom backbone of porcine pepsin. The view is approximately down the intramolecular 2-fold axis. The active site carboxylate side chains of Asp-32 and Asp-215 are shown in red. (**b**) Superposition of the C_α atoms of the revised pepsin structure (black lines) with those of the previously reported model (red lines) equivalenced over 326 residues. The orientation is rotated 90° about the vertical axis with respect to (**a**) to highlight the structural differences.

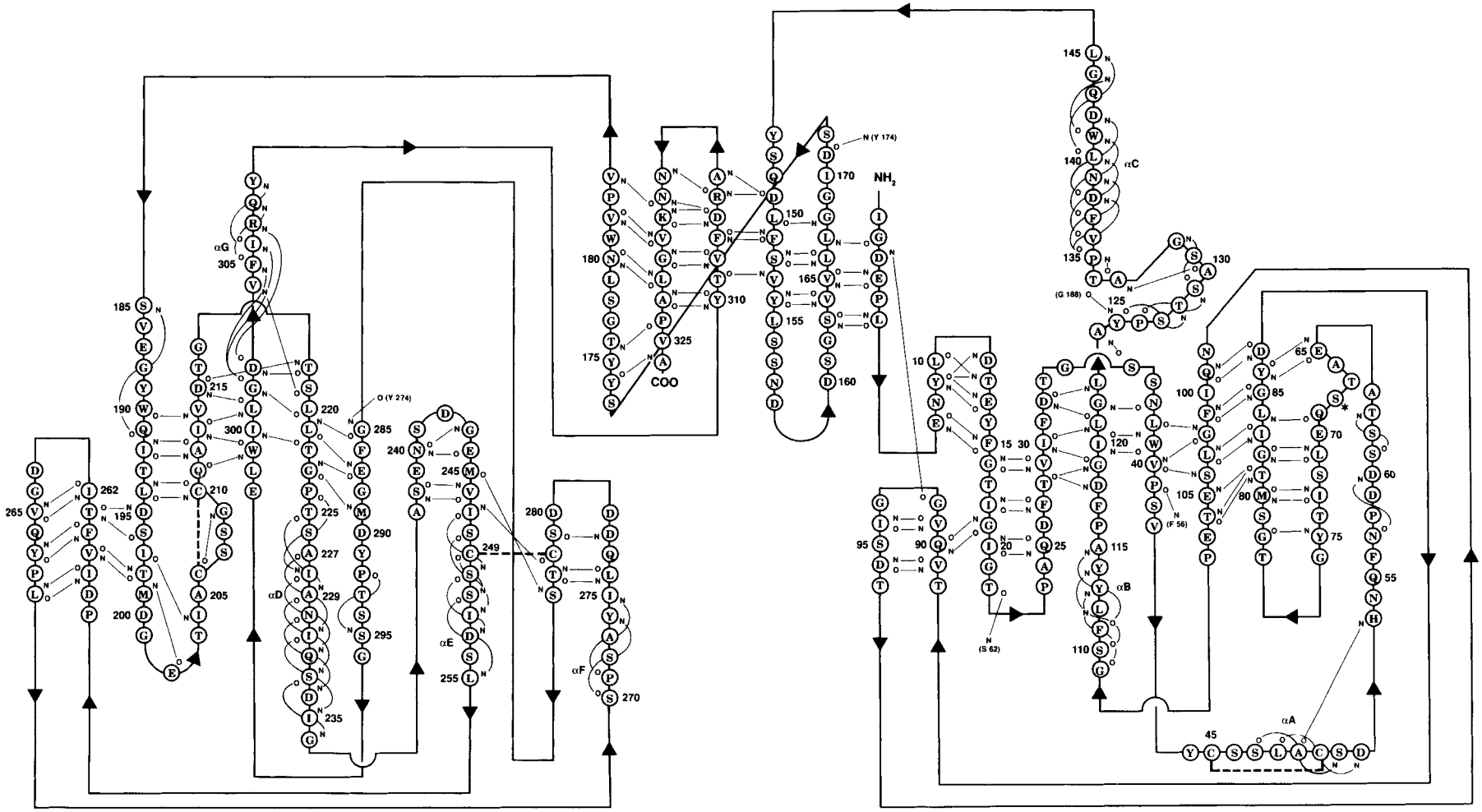

Fig. 5. Planar representation of the secondary structure of porcine pepsin. An asterisk indicates the phosphoserine at position 68. Disulfide bonds are represented by thick dashed lines.

TABLE II. β-Sheets in Pepsin*

β-sheet	Strand†	Residues	Description
I	βA1 (a_1)	2–5	Interdomain sheet; antiparallel
	βJ (j)	164–167	
	βI (l)	150–154	
	βV (g)	309–314	
	βX (r)	319–324	
	βK (k)	180–183	
II	βB (b)	15–20	Antiparallel
	βC1 (c)	25–28	
	βF3 (f_2)	87–91	
	βG1 (f_3)	94–97	
III	βL2 (f_3)	196–198	Antiparallel
	βM1 (g)	202–203	
	βQ (i)	258–262	
	βR (j)	265–269	
IV	βE3 (e_1)	70–74	"Flap" region; antiparallel
	βF1 (f)	79–83	
	βG3 (g_1)	105–106	
V	βO (n_1)	238–240	Two pairs of short β-strands linked by a disulfide formed from Cys-249–Cys-282
	βP (n_2)	245–247	
	βS (n_5)	275–277	
	βT1 (n_6)	282–284	
VI	βL1 (l)	191–195	Strands L1, M2, and N form the ψ2 sheet; mixed parallel and antiparallel
	βM2 (m_1)	210–214	
	βN (n)	221–223	
	βT2 (o)	286–288	
	βU (p_2)	298–301	
VII	βC2 (c_1)	29–32	Strands C2, D, and H form the ψ1 sheet; mixed parallel and antiparallel
	βD (d)	38–41	
	βH (h_1)	119–122	
	βE2	65–69	
	βF2	84–86	
	βG2	98–104	

*The algorithm of Kabsch and Sander[37] was used to assign limits for the secondary structural elements of pepsin.
†Aspartic proteinase strand designations used by Tang et al.[2] are listed in lower case letters in parentheses.

TABLE III. Helical Segments in Pepsin

Helix*	Residues
αA	Leu-48– Asp-52
αB	Ser-110– Tyr-114
αC (h_2)	Val-136– Asp-142
αD	Thr-225– Ile-235
αE	Cys-249– Leu-255
αF	Pro-271– Tyr-274
αG (p_5)	Asp-303– Ile-306

*Helix designations used by Tang et al.[2] for the fungal aspartic proteinases are provided in lower case in parentheses.

strengthened by recent structural studies of two retroviral proteases which demonstrated that these enzymes are structurally related to the eukaryotic aspartic proteinases, but exist as dimers of identical subunits which are each only 99–125 residues in length.[5–7]

Structural Comparisons With Fungal Aspartic Proteinases

We compared the refined structure of porcine pepsin with that of rhizopuspepsin, penicillopepsin, and endothiapepsin (Table VI and Figure 8).[8–10] Least-squares superposition of the 325 C_α atoms of rhizopuspepsin onto the 326 C_α atoms of pepsin resulted in 206 matching pairs with an overall rms deviation of 0.86 Å (Fig. 8 and Table VI). The minimization procedure superposed very well β-sheets I, II, and IV (rms = 0.78 Å for 71 C_α pairs) which comprise the N domain plus the last two β-strands of the C domain. The structures of the ψ_1 and ψ_2 β-sheets which form the active sites are well conserved (rms = 0.88 Å for 25 C_α pairs). In addition, β-sheet VI also superposed well (rms = 1.07 Å for 19 C_α pairs). On the other hand, β-sheets III and V, as well as helix αD, are displaced relative to the analogous structures in rhizopuspepsin (rms = 1.72 Å for only 4 equivalenced C_α pairs out of a total of 47 possible C_αs). Comparison of pepsin with penicillopepsin and endothiapepsin gave results similar to those obtained for the rhizopuspepsin/pepsin analysis. For penicillopepsin, the number of C_α pairwise equivalences was 179 with an rms of 0.83 Å. Endothiapepsin gave a total of 166 equivalenced C_α atom pairs with an rms of 0.82 Å. Overall, the results indicated that pepsin is less similar to the fungal proteinases than the latter are to each other.[8–10] The degree of struc-

TABLE IV. Polar Interactions Between Side Chains for Pepsin

Residue	Atom	Atom	Residue	Distance (Å)
Asn-8	ND2	OG	Ser-156	2.9
Asp-11	OD2	NH_2	Arg-307	3.3
Tyr-14	OH	NH_2	Arg-307	3.4
Thr-22	OG1	OG1	Thr-63	3.0
Gln-25	NE2	OG	Ser-61	3.0
Thr-28	OG1	OD2	Asp-118	3.1
Asp-32	OD1	OG	Ser-35	2.8
Thr-33	OG1	OG1	Thr-216	3.0
Asn-37	ND2	OG	Ser-129	3.2
Asp-54	OD1	OD2	Asp-118	2.8
Thr-63	OG1	OD2	Asp-87	2.7
Thr-81	OG1	OG1	Thr-106	2.9
Thr-97	OG1	OE1	Gln-143	2.8
Ser-104	OG	OE	Glu-107	2.7
Tyr-125	OH	OD1	Asn-318	3.0
Asp-138	OD2	NH2	Arg-315	2.8
Asn-139	ND2	OE1	Gln-143	2.9
Ser-152	OG	OH	Tyr-310	2.7
Ser-156	OG	OG	Ser-163	2.9
Asp-171	OD1	OG	Ser-173	2.8
Asn-180	OD1	OH	Tyr-267	3.1
Gln-187	OE2	ND2	Asn-318	2.9
Thr-198	OG1	OG1	Thr-203	2.6
Thr-255	OG1	OD1	Asp-290	3.0
Gln-232	OE1	OG	Ser-284	3.2
Ser-248	OG	OG	Ser-251	2.9
Asn-317	OD1	NZ	Lys-319	2.9
Asp-290	OD2	OE2	Glu-297	2.8

tural homology (number of equivalenced residues) between pepsin and the individual fungal enzymes follows the order rhizopuspepsin > penicillopepsin > endothiapepsin. Based upon this superposition analysis, a new sequence alignment for pepsin with the fungal aspartic proteinases was derived (Fig. 9).

Structural differences generally occurred in the loops and helices, and were most pronounced where there are gaps or insertions in the aligned amino acid sequences. For example, an insertion in pepsin of two residues after position 142 extends helix C by nearly one turn relative to the analogous helices in penicillopepsin and endothiapepsin (Fig. 8). An additional two residue insertion with respect to endothiapepsin and penicillopepsin after residue 251 results in the presence of helix αE in the carboxy domain of pepsin. Other examples include loops in pepsin formed by the additional disulfide Cys-206 and Cys-210 and by the relative insertion of 4–5 residues after position 291. The latter results in a large loop which extends in towards the active site cleft in pepsin and is partially disordered (see above). Sequence alignments[38] indicate that human and mouse renins, human and porcine cathepsins D, and the yeast aspartic proteinase A contain a comparable insertion in this region, whereas, bovine chymosin and embryonic chicken pepsin do not. Thus, the presence of this insertion is not a general feature of mammalian aspartic proteinases, and its structural significance remains unclear.

While many structural differences could be ascribed to insertions or deletions, large structural changes were also observed for several segments of identical length between pepsin and the fungal enzymes. A particularly striking example is the region encompassing Cys-45–Asp-52 in pepsin (rhizopuspepsin has a single deletion at position 48 or 49). This segment includes the N-domain disulfide bond (Cys–45–Cys-50), which is not present in two of the three fungal enzyme structures, and also forms the one-turn helix, αA. The analogous region in the fungal enzymes adopts various turn conformations. In the C domain, many poor correspondences can be observed between segments of conserved length (Fig. 8). However, as will be seen later (see Mobile Subdomain section), these apparent structural differences are largely due to positional shifts of similar structural elements (e.g., compare the positions of αD helices, Fig. 8) rather than to major conformational differences of the elements themselves.

Active Site Comparisons

The framework for the active site of the aspartic proteinase family is provided by the ψ loops in both the amino (ψ_1) and the carboxy (ψ_2) domains. Within these loops, the segments consisting of residues Ile-30–Ser-39, Ile-120–Ala-124, Ile-213–Thr-218, and Ala-301–Thr-305 have been used to compare the active sites of the different enzymes. Least-squares superposition of these four segments of the pepsin structure with the corresponding segments of the fungal enzymes gave rms differences ranging from

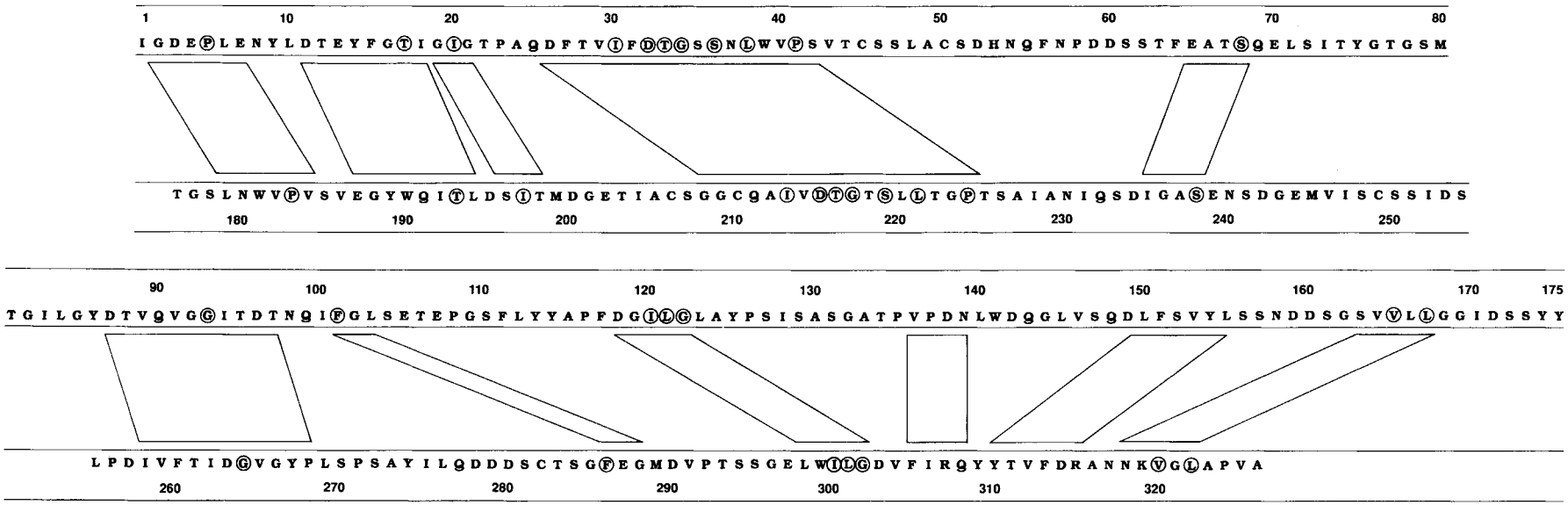

Fig. 6. Schematic representation of the alignment of the N and C domains of porcine pepsin. This alignment is based on a least-squares superposition of the C_α carbons. Topologically equivalent regions, defined as containing two or more consecutive, aligned C_α atom pairs, are bounded by parallelograms. Identical residues are circled.

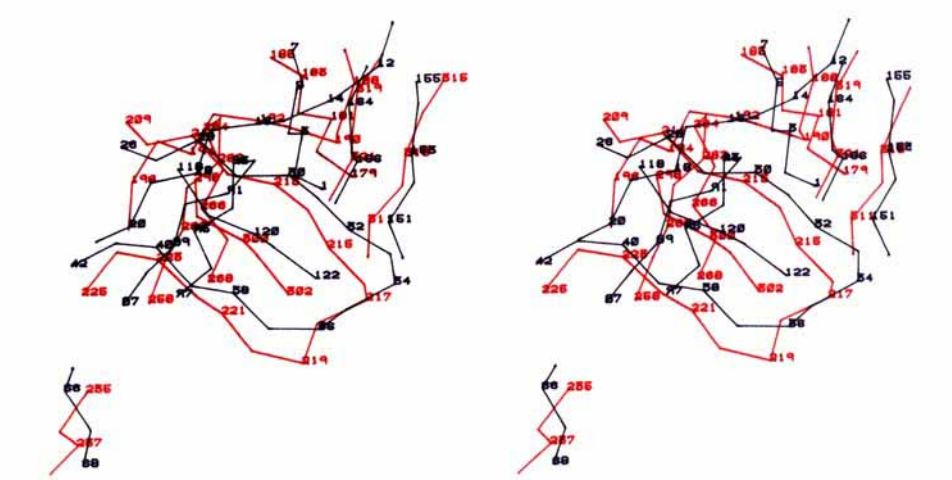

Fig. 7. Stereo diagram showing the superposition of the topologically equivalent regions in the N (black lines) and C (red lines) domains for porcine pepsin found by least-squares superposition analysis.

TABLE V. Interdomain Symmetry of Aspartic Proteinases

	No. equivalent C_α atoms	rms (Å)	Rotation angle (deg)
Rhizopuspepsin	77	2.4	173
Endothiapepsin	68	2.1	171
Penicillopepsin	70	1.9	177
Porcine pepsin	70	2.4	174

0.39 to 0.41 Å for the 88 main chain atom pairs (Table VII, Fig. 10). The active site structures of the fungal proteinases have a slightly better internal agreement with an average rms deviation of 0.24 Å. These values are similar to the ones found by Sielecki et al.[11] who compared recombinant human renin with the fungal proteinases.

Detailed comparisons between the active site structures (including solvent molecules) of pepsin and the fungal enzymes should be interpreted cautiously in view of the lower resolution of the pepsin structure. Nonetheless, certain features of the active site of pepsin are probably different from the fungal enzymes. The carboxyl groups of the active site aspartates are still essentially coplanar (rms deviation from the plane = 0.04 Å; largest deviation = 0.16 Å for Asp-32–OD2) and they are approximately the same distance apart (Asp-32 OD2–OD1 Asp-215 distance = 3.15 Å) as in the fungal enzymes (3.11 Å and 2.87 Å for rhizopuspepsin and penicillopepsin, respectively). This suggests that they are also hydrogen bonded in the mammalian enzyme. The orientation of the carboxyl group of Asp-215 is slightly different in pepsin versus the fungal enzymes due to a relative rotation about the χ^1 dihedral angle of approximately 20° toward the main chain of the ψ_1 loop (Fig. 10). The fungal enzymes were all crystallized at pH levels in which at least one of the two active site aspartates should be fully ionized. However, the pepsin crystals were grown at pH 2.0; it is likely that, on average, about 1.5 carboxylates will be protonated under these conditions. The differences in ionization state could account for the differences in the configuration of the aspartates.

The current model has preliminary positions for 206 water molecules with temperature factors ranging from 3.0 to 70 Å^2. Wat-596 is among the ones having the lowest B-factor (B = 6 Å^2) and its position is nearly identical to the internal water molecule corresponding to Wat-502 in rhizopus and Wat-4 in penicillopepsin. This water forms a hydrogen-bonding bridge between the N and C domain ψ loops (Thr-33 O–Wat-596–O Val-214) (Fig. 11a). Several other putative waters were found at positions that correspond to the two internal water clusters of the fungal enzymes. Electron density in the pepsin map is shown for one of these waters, Wat-757 (which corresponds to Wat-46 in penicillopepsin) (Fig. 11b). Even though this peak was refined as a water molecule, the shape of the electron density strongly suggests that an ethanol molecule occupies this position in the crystal structure of pepsin. This ethanol is located approximately 7 Å from the active site aspartates and is displaced by the binding of renin inhibitors to pepsin (Abad-Zapatero and Erickson, unpublished data). Based on the elongated shapes of many of the peaks in the final electron density map, as well as on the chemical environment of the maxima, it is estimated that as many as 10% of the putative waters are probably ethanol molecules.

In the latter stages of refinement for pepsin, a weak peak (0.5 σ) appeared in the electron density map at a position approximately equidistant from the two carboxylates. This peak was interpreted as a water molecule (Wat-418, Fig. 11c) and was refined

TABLE VI. Homologies Between Pepsin and the Fungal Aspartic Proteinases

Pepsin vs.	N-domain		C-domain		Overall		Overall sequence homology (%)
	#C_α pairs	rms (Å)	#C_α pairs	rms (Å)	#C_α pairs	rms (Å)	
Rhizopuspepsin	140	0.75	95	0.98	206	0.86	34
Penicillopepsin	107	0.72	71	0.78	179	0.83	30
Endothiapepsin	128	0.85	64	1.03	166	0.82	35

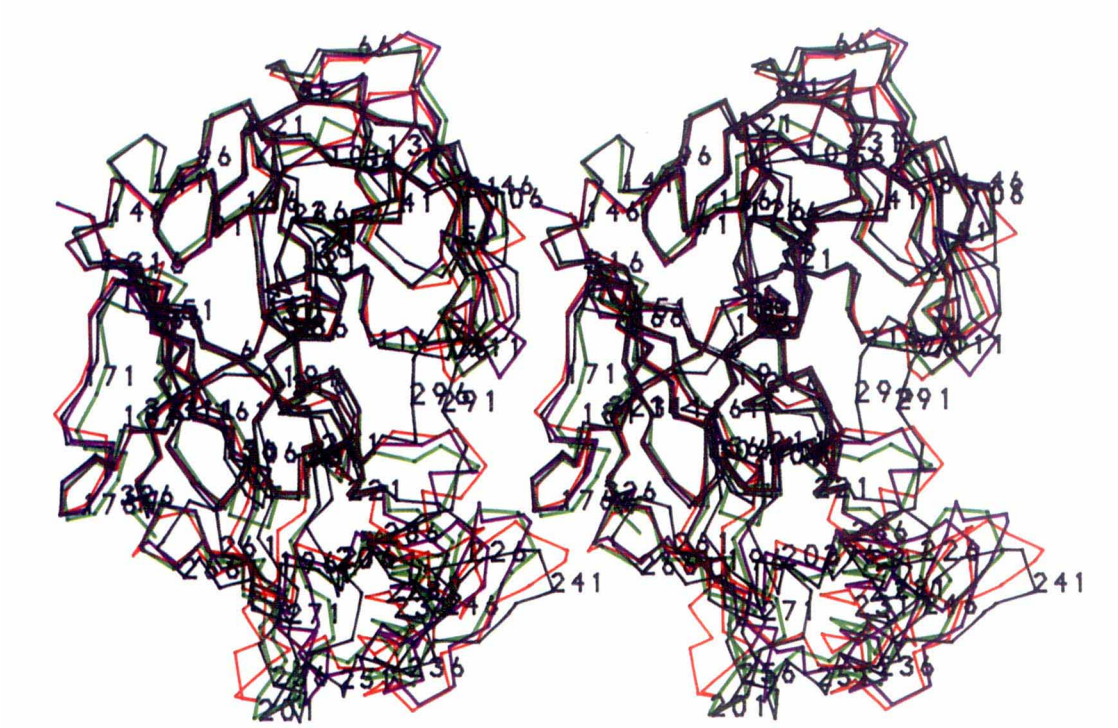

Fig. 8. Stereo diagram of the superposition of the C_α backbones of pepsin (black), rhizopuspepsin (green), penicillopepsin (violet), and endothiapepsin (orange). The amino acid numbering corresponds to pepsin. The figure presents the results of the overall superposition based on the method of Rossmann and Argos[24] as described in the experimental section.

to a final position which was 2.95 Å and 3.1 Å distant from Asp-32 OD2 and Asp-215 OD2, respectively. The fungal proteinases all contain a conserved water molecule in a nearly identical location to that observed for pepsin. The shape of the electron density for this peak is inconsistent with an ethanol molecule, which was previously identified as occupying this site in the earlier pepsin structure.[13] There are three other closely spaced electron density maxima ($\geq$ 1.0 σ) in the immediate vicinity of the active site with weak electron density bridges between them. The peak closest to Asp-32 OD1 (2.7 Å) corresponds to the refined position of Wat-752 with two other peaks (Wat-756 and Wat-452) within 3.0 Å of the first one. The latter peak is elongated and is surrounded by three hydrophobic side chains of Ile-30, Ile-120, and Tyr-75 of the "flap." It is conceivable that these three peaks correspond to the presence of a weakly bound ethanol molecule with two alternative positions for the OH group (Fig. 11c); in both conformers, the methyl group will be nestled against the hydrophobic pocket. Note that Ile-30 in pepsin replaces polar residues (Asp, Asn) in this position in the fungal enzymes and thus possibly enhances the hydrophobic character of this region near the active site.

The weak inhibition of porcine pepsin by ethanol[30] can be explained by our results as being due to two concomitant factors: (1) steric hindrance of the substrate due to the specific and nonspecific binding of ethanol molecules near the active site; (2) the presence of an ethanol molecule that may directly hydrogen bond to Asp-32 (or to Wat-418), and may compete, albeit weakly, with the substrate for the active site. Thus, the inhibition should appear kinetically as mixed competitive–noncompetitive,

```
PNO      1        10          19          29          39          49
END     STGSATTTPID SLDDAYITPVQ IGTPAQTLNL DFDTGSSDLW VFSSETTASE
PEN    AASGVATNTPTA N-DEEYITPVT IG--GTTLNL NFDTGSADLW VFSTELPASQ
RHI    AGVGTVPMTDYG N-DIEYYGQVT IGTPGKKFNL DFDTGSSDLW IASTLCTN--
PEP      IGDEPLENYL --DTEYFGTIG IGTPAQDFTV IFDTGSSNLW VPSVYCSSLA

PNO             62          71          80          90          100
END  V-DGQTIYTPSKST TAKLLSGATW SISYGDGSSS SGDVYTDTVS VGGLTVTGQA
PEN  Q-SGHSVYNPSAT- -GKELSGYTW SISYGDGSSA SGNVFTDSVT VGGVTAHGQA
RHI  CGSGQTKYDPNQSS TYQAD-GRTW SISYGDGSSA SGILAKDNVN LGGLLIKGQT
Por  C-SDHNQFNPDDSS TFEAT-SQEL SITYG-TGSM TGILGYDTVQ VGGISDTNQI

PNO            110         119         129         136         146
END      VESAKKVS-S SFTEDSTIDG LLGLAFSTLN TVSPTQQKTF FDNAKA--SL
PEN      VQAAQQIS-A QFQQDTNNDG LLGLAFSSIN TVQPQSQTTF FDTVKS--SL
RHI      IELAKREA-A SFASG-PNDG LLGLGFDTIT TV--RGVKTP MDNLISQGLI
PEP      FGLSETEPGS FLYYA-PFDG ILGLAYPSIS AS---GATPV FDNLWDQGLV

PNO            156         164         174         184         193
END      DSPVFTADLG Y---H-A-PGT YNFGFIDTTA YTGSITYTAV STKQGFWEWT
PEN      AQPLFAVALK H---Q-Q-PGV YDFGFIDSSK YTGSLTYTGV DNSQGFWSFN
RHI      SRPIFGVYLG KA-KNGG-GGE YIFGGYDSTK FKGSLTTVPI DNSRGWWGIT
PEP      SQDLFSVYLS SN--D-DSGSV VLLGGIDSSY YTGSLNWVPV SVE-GYWQIT

PNO            203         213         223         233         241
END      STGYAVG-SG TFKSTSIDGI ADTGTTLLYL PATVVSAYWA QVSGAKSSSS
PEN      VDSYTAGSQS ---GDGFSGI ADTGTTLLLL DDSVVSQYYS QVSGAQQDSN
RHI      VDRATVGTST VA-S-SFDGI LDTGTTLLIL PNNIAASVAR AYG-ASDNG-
PEP      LDSITMDGET IACSGGCQAI VDTGTSLLTG PTSAIANIQS DIG-ASENS-

PNO            251         261         271         278         288
END      VGGYVFPCSA --TLPSFTFG VGSARIVIPG DYIDFGPIST GSSSCFGGIQS
PEN      AGGYVFDCST --NLPDFSVS ISGYTATVPG SLINYGPSGD G-STCLGGIQS
RHI      DGTYTISCD- TSAFKPLVFS INGASFQVSP DSLVFEE-F- Q-GQCIAGFGY
PEP      DGEMVISCSS IDSLPDIVFT INGVQYPLSP SAYIL-Q-D- D-DSCTSGFEG

PNO            298         308         317      326
END      SAG-IG---I NIFGDVALKA AFVVFNGATT PTLGFASK
PEN      NSG-IG---F SIFGDIFLKS QYVVFDSD-G PQLGFAPQA
RHI      GN--WG---F AIIGDTFLKN NYVVFNQG-V PEVQIAPVAE
PEP      MDVPTSSGEL WILGDVFIRQ YYTVFDRA-N NKVGLAPVA
```

Fig. 9. Sequence alignments for porcine pepsin and the fungal aspartic proteinases based upon the three-dimensional superpositions of their C_α coordinates. Abbreviations: PNO, pepsin numbering; END, endothiapepsin; PEN, penicillopepsin; RHI, rhizopuspepsin; PEP, porcine pepsin. Amino acid sequences were taken from Ref. 38.

TABLE VII. Active Site Superposition Among Aspartic Proteinases*

	Pepsin	Endothia	Rhizopus	Penicillum
Pepsin	—	0.421	0.39	0.40
Endothiapepsin	0.37	—	0.28	0.22
Rhizopuspepsin	0.35	0.26	—	0.24
Penicillopepsin	0.39	0.21	0.21	—

*The numbers above the diagonal are the rms deviations, in Å, between 88 pairs of main chain atoms (N, C_α, C, O) for the corresponding enzymes. The numbers below the diagonal correspond to the superposition between C_α atoms only (22 pairs).

and is consistent with recent observations (E. Matayoshi, unpublished data).

Evidence for a Flexible Subdomain

Suguna et al.,[10] in their structural comparison of rhizopuspepsin and penicillopepsin, found that the N domains were more similar (1.08 Å, 162 C_α pairs) than the C domains (1.37 Å, 136 C_α pairs) and that the structural differences for each domain were distributed approximately uniformly. This observation is consistent with the fact that the C domains, and in

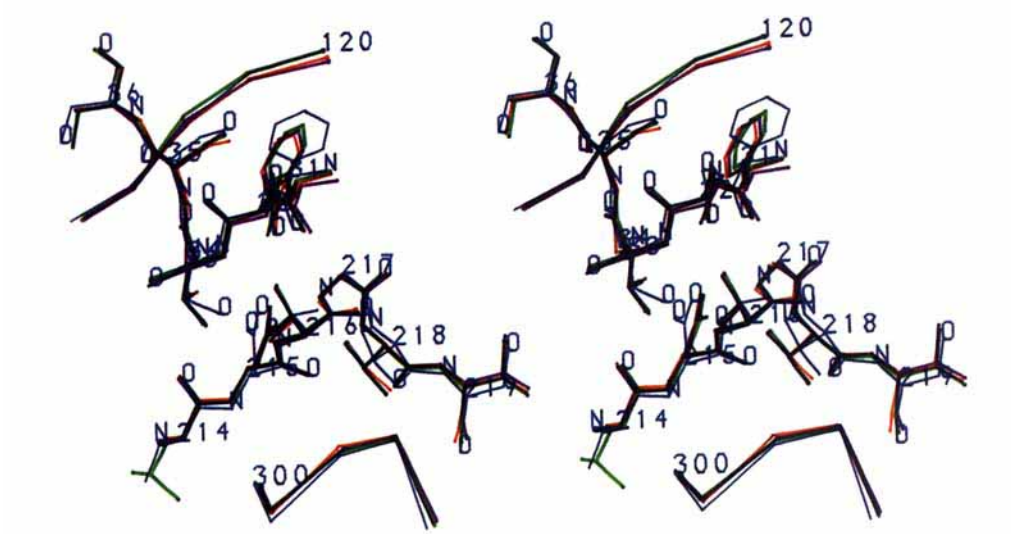

Fig. 10. Superposition of the active site residues of porcine pepsin (blue) on the corresponding residues for the fungal aspartic proteinases. The same color code as in Figure 8 was used for the fungal enzymes. All the atoms are shown for the most strictly conserved amino acid residues (31–36; 214–219); only C_α atoms are shown for the residues of the structurally conserved psi loops (120–124; 300–304).

particular residues 223–299, exhibit less sequence homology for the aspartic proteinases (Fig. 9). We also observed that the largest structural differences between pepsin and the fungal proteinases occur among the C domains (Figs. 8 and 12a). In fact, up to twice as many residues can be superposed between the N domains as between the C domains (Table VI). However, in contrast to the case for the rhizopuspepsin/penicillopepsin comparison,[10] the structural differences between pepsin and the fungal enzymes were not uniformly distributed, but were instead largely localized to the region of low sequence homology (residues 223–299) within the C domains. For example, using the overall least-squares superposition of penicillopepsin onto pepsin, only two short stretches of 7 and 4 C_α atom pairs superimposed between residues 222 and 298. Similarly, only 6 C_α pairs superimposed for the rhizopuspepsin/pepsin and none for the endothiapepsin/pepsin superpositions. Thus, although the C domains of the mammalian and fungal enzymes possess the same folding architecture and very similar secondary structures, they do not correspond spatially when subjected to overall least-squares superposition analysis. One possible explanation for this result is that a concerted rigid body movement of the C domain relative to the N domain could account for the observed deviations. However, this cannot be the case since approximately one-third of the residues in the C domain (residues 177–190, 211–221, and 296–326) align reasonably well along with the majority of the residues in the N domain using the overall superposition. An alternative explanation is that there exists, within the C domain, a subdomain that can assume various conformations for the different enzymes. The latter hypothesis was tested by performing least squares superpositions for the segment comprising residues 222–298 of pepsin with the corresponding segments for the fungal proteinases. This analysis revealed that these segments could indeed be well aligned if they are allowed to move independently of the rest of the molecule during the superposition analysis (Table VIIIA and Fig. 12b). Indeed, 64–72% (up to 50 C_α atom pairs) of the residues in these segments for the fungal enzymes aligned to better than 1.7 Å rms with the corresponding segment in pepsin. This was a dramatic improvement over the alignment obtained from the overall superposition analysis, and confirmed the ex-

Fig. 11. Solvent structure in the active site vicinity of pepsin. Crosses indicate refined positions for the oxygen atoms of putative water molecules. (**a**) Stereoscopic view of the density associated with the refined position of Wat-596 and its immediate environment. (**b**) Stereoscopic view of the density corresponding to the refined position of Wat-757. The shape of the maxima and the environment in its vicinity strongly suggests that the maxima corresponds to an ethanol molecule; such a molecule has been modeled to fit the density but not refined. The protein ligands to the hydroxyl group of ethanol are OD2 Asp-303 and O Thr-12. O418 marks the position of the water molecule approximately equidistant from the active site aspartates. (**c**) Stereoscopic view of the solvent maxima in the immediate vicinity of the active site aspartates. Dashed lines depict possible interactions among the different atoms (≤3.7 Å). The water molecule shared by Asp-32 and Asp-215 is 418 and corresponds to the solvent molecule conserved among all aspartic proteinase whose structure is known. The three solvent peaks (752, 452, and 756) were refined as three water molecules. Thin lines inside these peaks suggest an alternative interpretation of the maxima as an ethanol molecule having two different positions for its hydroxyl group. The methyl group would correspond to O452 and would be in the hydrophobic pocket formed by Ile-30, Ile-120, and the phenyl ring of Try-75.

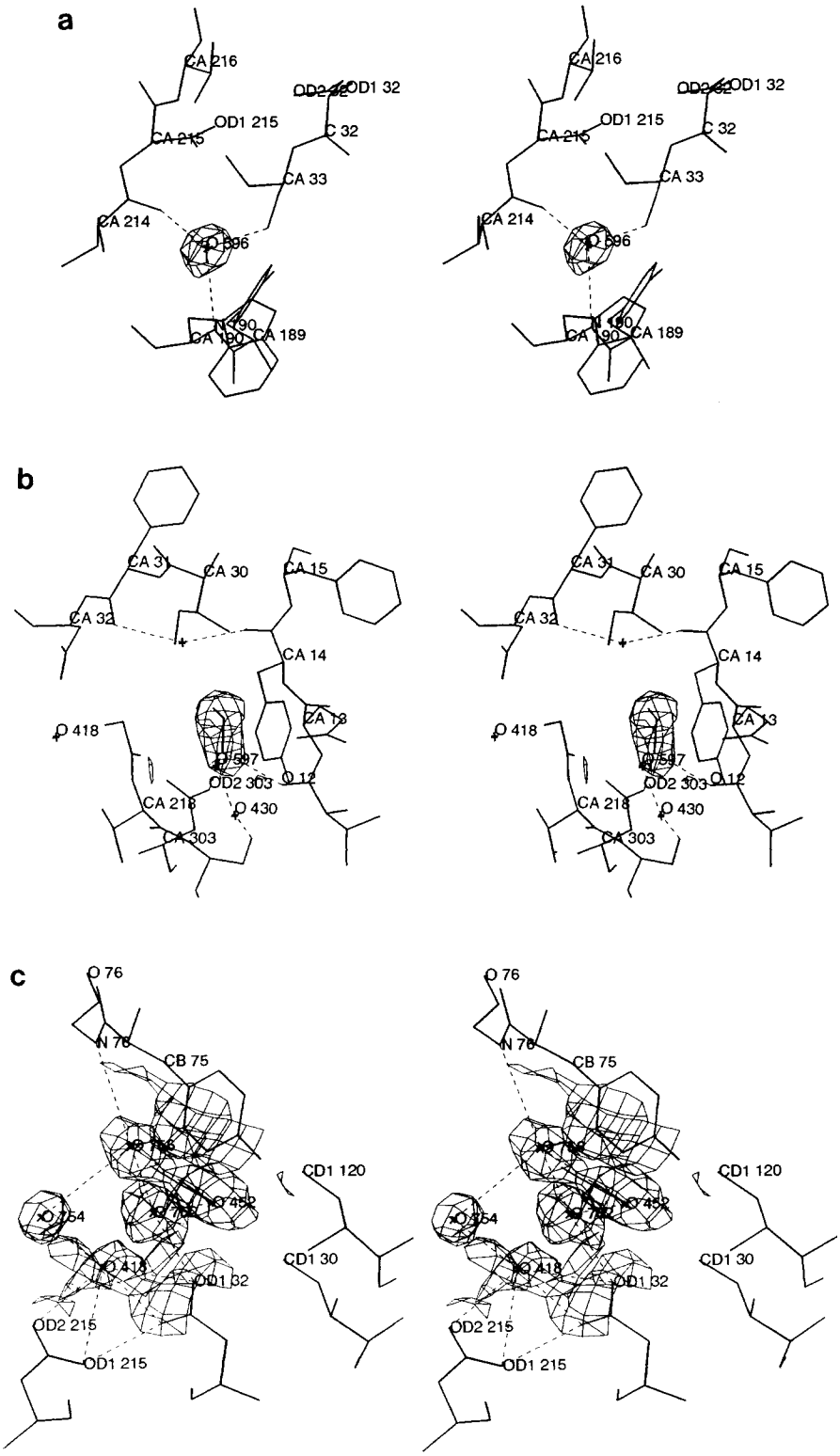

Fig. 11. Legend appears on page 76.

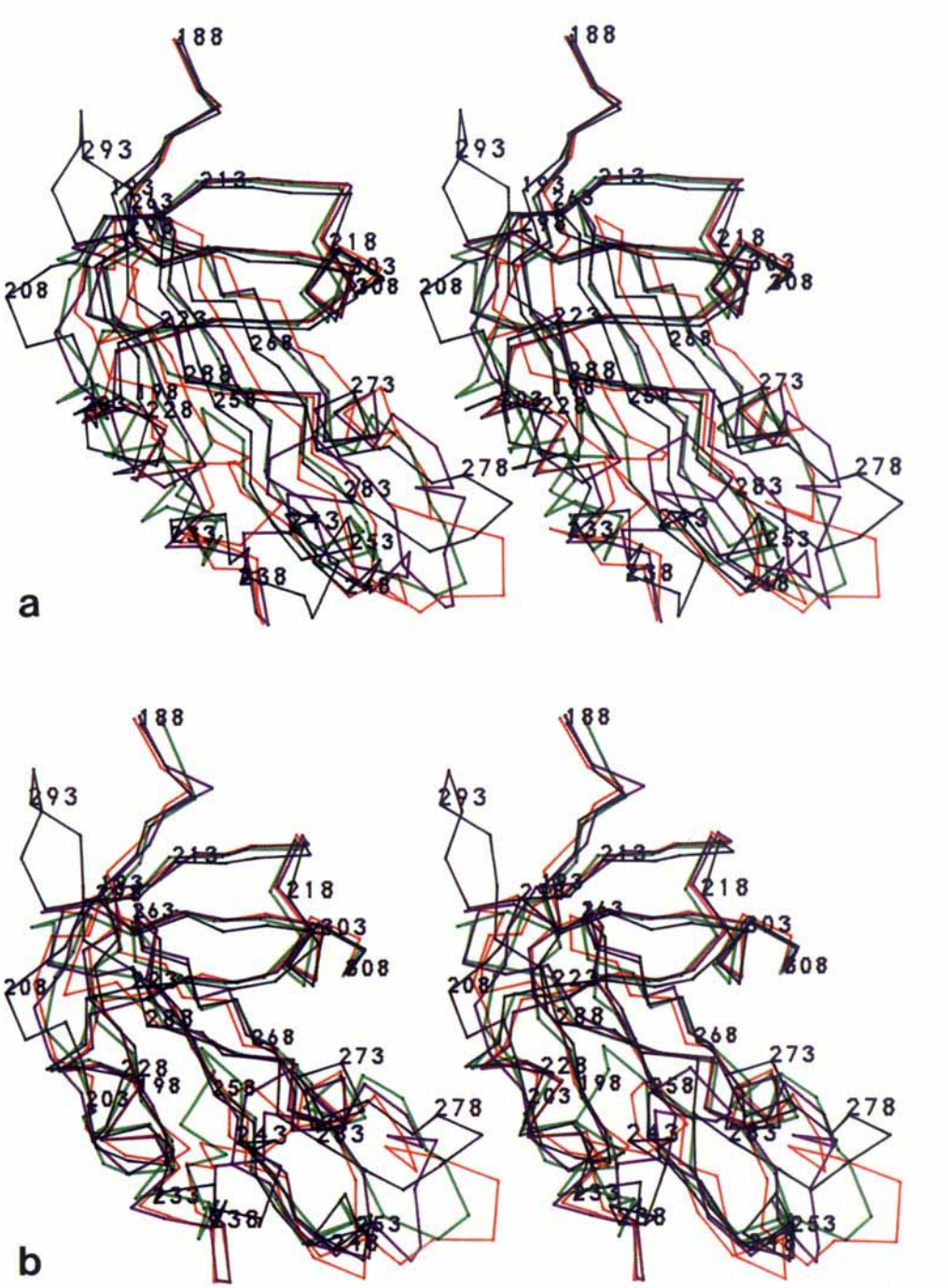

Fig. 12. Superposition of the flexible subdomains of the aspartic proteinases. (**a**) Relative conformations obtained after overall superposition of the fungal enzymes onto pepsin. (**b**) The resulting relative conformations of the subdomains after application of rigid body transformations that were derived by superposing the subdomain segments listed in Table VIIIA. These transformations were applied to the fungal enzyme subdomain segments corresponding to residues 192–212 and 223–299 (pepsin numbering), keeping the intervening ψ_2-loop residues fixed. This attempt to simulate the hinge motion resulted in slight distortions of the polypeptide backbone at the points of connection between the subdomain and sheet VI. Endothiapepsin (yellow), rhizopuspepsin (green), and penicillopepsin (violet) in relation to pepsin (blue). The pepsin numbering is indicated.

istence of a structurally distinct subdomain in these enzymes which had not been previously described.

Rigid body analysis of the subdomain displacements relative to pepsin indicated that the subdomains were related by a combination of a rotation, ranging from 11.2–21.0°, and an overall translation of 3.2–5.9 Å (Table VIIIA). The subdomain of endothiapepsin exhibited by far the largest displacement

TABLE VIII. Relative Subdomain Displacements for the Aspartic Proteinases*

	Residues	rms (Å)	#C_α pairs	Rotation (deg)	Translation (Å)			
					X	*Y*	*Z*	Overall
			A. Porcine Pepsin vs.					
Penicillopepsin	221–295	1.68	46	12.2	2.9	−0.9	1.7	3.9
Endothiapepsin	222–299	1.59	44	21.0	4.9	−1.8	2.7	5.9
Rhizopuspepsin	226–297	1.12	50	11.2	1.8	0.0	2.6	3.2
			B. Endothiapepsin vs.					
Rhizopuspepsin	222–297	1.20	46	8.4	−2.2	0.4	−1.4	2.6
Penicillopepsin	221–295	0.63	49	14.5	−1.4	−2.2	−2.9	3.9
			C. Rhizopuspepsin vs.					
Penicillopepsin	221–295	1.17	43	7.5	0.9	−2.3	−1.6	2.9

*Residue members refer to the sequences of the fungal enzymes.

in relation to pepsin. However, the final superposition of the endothiapepsin/pepsin subdomains was slightly better (1.59 Å) than that for the penicillopepsin/pepsin subdomains (1.68 Å), even though the latter pair deviated less in the overall superposition. These observations indicate that overall topological similarity does not necessarily correlate with the magnitude of the rigid body displacements for the subdomains. Analysis of the subdomain displacements observed between pepsin and the fungal enzymes was further extended to the fungal enzymes, themselves (Table VIIIB and C). As for the case with the pepsin/fungal enzyme comparisons, the structural homology for these segments improved markedly when they were allowed to move as a rigid body relative to the rest of the molecule. Relative subdomain displacements of up to 14.5° rotation and 3.9 Å translation were observed, with the largest displacements occuring between penicillopepsin and endothiapepsin.

The subdomain displacements observed for the aspartic proteinases suggests a structural polymorphism which had not been previously described for this family of enzymes. The results of the subdomain superposition analysis could be explained by assuming that these segments possess flexibility and may exhibit some limited range of motion relative to the rest of the molecule. Flexibility for the subdomain of pepsin is supported by the significantly larger overall temperature factors for this region ($B > 35$ Å^2 for residues 230–295). Moreover, the analogous regions in rhizopuspepsin[10] and penicillopepsin[8] also were reported to exhibit above average *B* factors. Further evidence for the adjustable nature of this subdomain in pepsin is suggested by our crystallographic studies of pepsin/renin inhibitor complexes which indicate that the subdomain is displaced in the complex by approximately a 3.0° rotation and a 1.8 Å translation relative to the apoenzyme conformation (Abad-Zapatero and Erickson, unpublished data). In addition, a preliminary comparison of the structures of pepsin with that of porcine pepsinogen (Remington and Hartsuck, private communication) shows a significant subdomain displacement between the zymogen and the enzyme. Therefore, we conclude that this segment of aspartic proteinase structure may be a flexible subdomain whose potential range of motion needs to be more clearly delineated for each individual enzyme.

Subdomain Structure

The subdomain in pepsin is comprised mainly of the irregular sheets III and V plus helices αD, αE, and αF, and includes the only disulfide bond (Cys-249–Cys-282) which is strictly conserved in all known mammalian and fungal aspartic proteinases (see Figs. 3 and 4). The disposition of the subdomain with respect to the rest of the molecule is mediated through its connections to the five strands of sheet VI (strands L1, M2, U, N, and T2). This mixed β-sheet forms the active site ψ_2-loop, whose conformation is highly conserved among the enzymes, and connects the subdomain, through strands βL1 and βU, with the interdomain sheet I. The orientation of the plane of sheet VI varies for the different enzymes (see Fig. 8). Hence, sheet VI may serve as a flexible, multistrand hinge which allows for relative movement of the subdomain. Viewed from the outermost strand, βT_2, of sheet VI, the rotational component can be described as a clockwise hinge movement about an axis approximately passing through the C_a atoms of residues 192, 212, 299, 223, and 288 (Fig. 12). This movement is qualitatively and quantitatively different from that recently described for a similar segment in endothiapepsin.[31] In the latter case, the movement is apparently induced by inhibitor binding, and consists of a 4.1° rotation with little or no translation about an axis passing through residues 22 and 213. This axis is nearly perpendicular to the hinge axis we described for the subdomain motion, indicating that several types of subdomain motion may occur in these enzymes.

It is possible that the conformation of the subdomain is influenced by different crystalline environments. Although weak in macromolecular crystals, packing forces are known to affect the conformation of short, external loops involved in crystal contacts.[32–34] However, the size of the subdomains,

as well as the observed magnitudes of their relative displacements, percludes crystal packing as being the sole or even major determinant of the observed structural differences. Moreover, analysis of the crystal packing in pepsin shows that extensive regions of the subdomain (residues 222–235; 255–276; and 283–290) are not involved in crystal contacts. Thus, we suggest that the different structures resulting from the various orientations of this subdomain are not artifacts of crystal packing, but are partly a reflection of the segmental flexibility of this region, in which particular conformations have been selected in the various crystal forms of aspartic proteinase structures.

The functional role for the flexible subdomain is unclear, as are the factors that govern its conformation. We have observed that portions of the subdomain for pepsin can interact with inhibitors, and therefore probably contribute to the structural delineation of S′ subsites. It is possible that rigid body movements of the subdomain relative to the large domain may help the enzyme adjust to various substrate structures. Local regional flexibility of structures, such as the "flap," in and around the active site has been suggested to modulate substrate and inhibitor binding.[35] Perhaps the subdomain motion could provide coarse structural adjustments for accommodating widely different polypeptide sequences. The smaller scale, local movements could provide for fine tuning of the fit. This hypothesis suggests that the degree of mobility about the hinge region may be related to the range of substrate specificities which the enzyme can accommodate. Renin, for example, which has a narrow specificity, would be predicted to have very limited subdomain flexibility compared with pepsin. An additional role for subdomain flexibility may be in autoactivation. The activation of pepsinogen to pepsin requires major structural rearrangements in both the N and C domains.[36] It is tempting to speculate that the inability of prorenin to autoactivate, in contrast to pepsin and the fungal enzymes, may be partly a reflection of a more rigid hinge structure.

SUMMARY AND CONCLUSIONS

The revised structure of porcine pepsin confirms many of the observations reported earlier.[13] It provides further support to the notion that the aspartic proteinases may have evolved by gene duplication from an ancestral protein of approximately half the size. The structure also strengthens the evidence for the presence of a 4-fold repeat or supersecondary structure within this class of enzymes, although the evolutionary significance of the intradomain dyad has recently been questioned.[39] The present structure of the active site is not fully consistent with that reported earlier, in particular with respect to the disposition of the active site carboxylates and with the location of ethanol molecules.[13] The active site structures of pepsin and the fungal enzymes are largely conserved overall. The differences in the conformation of the carboxylate of Asp-215 in pepsin may be due to differences in the ionization states of the active site aspartates. The active site structure could also be influenced by the presence of ethanol molecules which replace several waters and alter the hydrogen-bonding network within the active site vicinity. Comparison of the current structure of pepsin with a higher pH crystal form without ethanol could provide insights into the roles of pH and ethanol on the active site structure.

Although the appearance and internal symmetry of the pepsin structure (and other aspartic proteinases) support its description as a bilobal entity of approximately equal parts, the discovery of a flexible subdomain suggests an alternative description of the structure. It may be useful to consider the molecule as consisting of two unequal domains. A large, relatively rigid domain comprises nearly two-thirds of the residues and includes the entire N domain as well as portions of the C domain. A smaller, flexible subdomain is localized within the C domain. The subdomain is connected to the large domain via β-sheet VI, and can be visualized as extending or hanging down from this sheet. We speculate that flexibility of the subdomain may play a role in mediating substrate binding, delineating substrate specificity, or possibly in the activation of zymogens for this family of enzymes. Elucidation of the limits and determining factors of segmental mobility of the subdomain is an exciting challenge that lies ahead for continued structural studies of this important family of enzymes.

ACKNOWLEDGMENTS

We wish to thank Dr. Ed Westbrook and Mary Deacon for assistance with the area detector data collection at Argonne National Laboratories. We especially thank Dr. Natalia Andreeva and her colleagues for providing the starting set of coordinates for pepsin which were used in this study. We gratefully acknowledge Dr. Tom Blundell for providing us with the endothiapepsin coordinates refined to 1.8 Å resolution. We thank Dr. Jonathan Greer for critically reading the manuscript, and for making several useful suggestions. In addition, we sincerely appreciate the assistance of Pat Collins and Cynthia Davis with the preparation of the illustrations and text, respectively.

REFERENCES

1. Tang, J., Wong, R.N.S. Evolution in the structure and function of aspartic proteases. J. Cell. Biochem. 33:53–63, 1987.
2. Tang, J., James, M.N.G., Hsu, I.N., Jenkins, J.A., Blundell, T.L. Structural evidence for gene duplication in the evolution of the acid proteases. Nature (London) 271:618–621, 1978.
3. Kostka, V. (ed.). "Aspartic Proteinases and Their Inhibitors." New York: Walter de Gruyter, 1985.

4. Pearl, L.N., Taylor, W.R. A structure model for the retroviral proteases. Nature (London) 329:351–354, 1987.
5. Miller, M., Jaskolski, M., Rao, J.K.M., Leis, J., Wlodawer, A. Crystal structure of a retroviral protease proves relationship to aspartic protease family. Nature (London) 337: 576, 1989.
6. Navia, M., Fitzgerald, P., McKeever, B.M., Leu, C-T., Heimbach, J.C., Herber, W.K., Sigal, I.S., Darke, P.L., Springer, J.P. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature (London) 337:615–619, 1989.
7. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J., Kent, S.B.H. Conserved folding in retroviral proteases: Crystal structure of a synthetic HIV-1 protease. Science 245:616–621, 1989.
8. James, M.N.G., Sielecki, A.R. Structure and refinement of penicillopepsin at 1.8 Å resolution. J. Mol. Biol. 163:299–361, 1983.
9. Blundell, T., Jenkins, J., Pearl, L., Sewell T. The high resolution structure of endothiapepsin. In: "Aspartic Proteinases and Their Inhibitors." Kostka, V. (ed.). New York: Walter de Gruyter, 1985:151–161.
10. Suguna, K., Bott, R.R., Padlan, E.A., Subramanian, E., Sheriff, S., Cohen, G.H., Davies, D.R. Structure and refinement at 1.8 Å resolution of the aspartic proteinase from rhizopus chinensis. J. Mol. Biol. 196:877–900, 1987.
11. Sielecki, A.R., Hayakawa, K., Fujinaga, M., Murphy, M.E.P., Fraser, M., Muir, A.K., Carilli, C.T., Lewicki, J.A., Baxter, J.D., James, M.N.G. Structure of recombinant human renin, a target for cardiovascular-active drugs, at 2.5Å resolution. Science 243:1346–1351, 1989.
12. Andreeva, N.S., Federov, A.A., Gustchina, A.E., Riskulov, R.R., Safro, M.G., Shutzkever, N.E. X-ray diffraction analysis of pepsin. 5. Conformation of the backbone of the enzyme. Mol. Biol. Engl. Transl. Mol. Biol. (Mosc.) 12:704–707, 1978.
13. Andreeva, N.S., Zdanov, A.S., Gustchina, A.E., Federov, A.A. Structure of ethanol-inhibited porcine pepsin at 2 Å resolution and binding of the methyl ester of phenylalanyldiiodotyrosine to the enzyme. J. Biol. Chem. 259:11353–11365, 1984.
14. Erickson, J., Abad-Zapatero, C., Rydel, T.J., Luly, J. The 2.5 Å crystal structure of a potent human renin inhibitor bound to porcine pepsin. In: "Fourteenth International Congress of Crystallography, Collected Abstracts." Freeman, H.C. (ed.). Perth, Australia: Lamb Printers, 1987: C-35.
15. Fesik, S.W., Luly, J.R., Erickson, J.W., Abad-Zapatero, C. Isotope-edited proton NMR study on the structure of a pepsin/renin inhibitor complex. Biochemistry 27:8297–8301, 1988.
16. Sham, H.L., Bolis, G., Stein, H.H., Fesik, S.W., Marcotte, P.A., Plattner, J.J., Rempel, C.A., Greer, J. Renin inhibitors. Design and synthesis of a new class of conformationally restricted analogues of angiotensinogen. J. Med. Chem. 31:289–295, 1988.
17. Bolis, G. Greer, J. Role of computer-aided molecular modeling in the design of novel inhibitors of renin. In: "Computer-Aided Drug Design. Methods and Applications." Perun, T.J., Propst, C.L. (es.). New York: Dekker, 1989: 297–326.
18. Howard, A.J., Gilliland, G.L., Finzel, B., Poulos, T.L., Ohlendorf, D.H., Salemme, F.R. The use of an imaging proportional counter in macromolecular crystallography. J. Appl. Crystallogr. 20:383–387, 1987.
19. Herzberg, O., Sussman, J.L. Protein model-building by the use of a constrained-restrained least-squares procedure. J. Appl. Crystallogr. 16:144–150, 1983.
20. Jones, T.A. In: "Computational Crystallography." Sayre, D. (ed). Oxford: Clarendon Press, 1982: 303–317.
21. Hendrickson, W.A., Konnert, J.H. In: "Biomolecular Structure, Conformation, Function, and Evolution," Vol. 1. Srinivasan, R. (ed.). Oxford: Pergamon Press, 1981: 43–57.
22. Brunger, A.T., Krukowski, A., Erickson, J.W. Slow-cooling protocols for crystallographic refinement by simulated annealing. Acta Cryst A46: in press, 1990.
23. Bernstein, F.C., Koetzle, T.F., Williams, G.F., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
24. Rossmann, M.G., Argos, P. A comparison of the heme binding pocket in globins and cytochrome b_5. J. Biol. Chem. 250:7525–7532, 1975.
25. Moravek, L., Kostka, V. Complete amino acid sequence of hog pepsin. FEBS Lett. 43:207–211, 1974.
26. Lin, X., Wong, R.N.S., Tang, J. Synthesis, purification, and active site mutagenesis of recombinant porcine pepsinogen. J. biol. Chem. 264:4482–4489, 1989.
27. Andreeva, N.S., Gustchina, A.E. On the supersecondary structure of acid proteases. Biochem. Biophys. Res. Commun. 87:32–42, 1979.
28. Richardson, J.S. The anatomy and taxonomy of protein structure. Adv. Protein Chem. 34:167–339, 1981.
29. Blundell, T.L., Sewell, B.T., McLachlan, A.D. Four-fold structural repeat in the acid proteases. Biochim. Biophys. Acta 580:24–31, 1979.
30. Antonov, V.K. New data on pepsin mechanism and specificity. Adv. Exp. Med. Biol. 95:179–198, 1977.
31. Sali, A., Veerapandian, B., Cooper, J.B., Foundling, S.I., Hoover, D.J., Blundell, T.L. 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, 1989.
32. Wang, D., Bode, W., Muser, R. Bovine chymotrypsinogen A. X-ray crystal structure analysis and refinement of a new crystal form at 1.8 Å resolution. J. Mol. Biol. 85:595–624, 1985.
33. Birktoft, J.J., Blow, D.M. Structure of crystalline α-chymotrypsin. 5. Atomic structure of tosyl-chymotrypsin at 2 Å resolution. J. Mol. Biol. 68:187–240, 1972.
34. Bode, W., Chen, Z., Bartels, K., Kutzbach, C., Schmidt-Kastner, G., Bartunik, H. Refined 2 Å X-ray crystal structure of porcine pancreatic Kallikrein A, a specific trypsin like serine proteinase. Crystallization, structure determination, crystallographic refinement, structure and its comparison with bovine trypsin J. Mol. Biol. 64:237–282.
35. James, M.N.G., Sielecki, A., Salituro, F., Rich, D.H., Hofmann, T. Conformational flexibility in the active sites of aspartyl proteinases revealed by a pepstatin fragment binding to penicillopepsin. Proc. Natl. Acad. Sci. U.S.A. 79:6137–6141, 1982.
36. James, M.G.M., Sielecki, A.R. Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8 Å resolution. Nature (London) 319:33–38, 1985.
37. Kabsch, W., Sander, C. Dictionary of protein structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2537, 1983.
38. Foltmann, B. Aspartic proteinases: Alignment of amino acid sequences. In: "Proceedings of the 18th Linderstrom-Lang Conference: Aspartic Proteinases." Foltmann, B. (ed.). Copenhagen: University of Copenhagen, 1988: 7–20.
39. Rao, J.K.M., Wlodawer, A. Is the pseudo-dyad in retroviral proteinase monomers structural or evolutionary? FEBS Lett., in press, 1990.
40. Peters, D., Peters, J. Quantum theory of the structure and bonding in proteins. 8. The alanine dipeptide. J. Mol. Struct. 2:107–123, 1981.