PROTEINS: Structure, Function, and Genetics 13:1–25 (1992)

The High-Resolution Crystal Structure of Porcine Pepsinogen

Jean A. Hartsuck,[1] Gerald Koelsch,[1] and S. James Remington[2]
[1]Protein Studies Program, Oklahoma Medical Research Foundation and Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, 73104, and [2]Institute of Molecular Biology, University of Oregon, Eugene, Oregon, 97403-1229

ABSTRACT **The structure of porcine pepsinogen at pH 6.1 has been refined to an *R*-factor of 0.173 for data extending to 1.65 Å. The final model contains 180 solvent molecules and lacks density for residues 157–161. The structure of this aspartic proteinase zymogen possesses many of the characteristics of pepsin, the mature enzyme. The secondary structure of the zymogen consists predominantly of β-sheet, with an approximate 2-fold axis of symmetry. The activation peptide packs into the active site cleft, and the N-terminus (1P–9P) occupies the position of the mature N-terminus (1–9). Thus changes upon activation include excision of the activation peptide and proper relocation of the mature N-terminus. The activation peptide or residues of the displaced mature N-terminus make specific interactions with the substrate binding subsites. The active site of pepsinogen is intact; thus the lack of activity of pepsinogen is not due to a deformation of the active site. Nine ion pairs in pepsinogen may be important in the advent of activation and involve the activation peptide or regions of the mature N-terminus which are relocated in the mature enzyme. The activation peptide–pepsin junction, 44P-1, is characterized by high thermal parameters and weak density, indicating a flexible structure which would be accessible to cleavage. Pepsinogen is an appropriate model for the structures of other zymogens in the aspartic proteinase family.**

Key words: aspartic proteinase zymogen, molecular replacement, structure–function, activation peptide, acid activation

INTRODUCTION

The aspartic proteinase family of enzymes (EC 3.4.23) shares a common catalytic apparatus composed of two aspartic acid side chains. In most cases, their maximal enzymic activity occurs at an acidic pH. The best known members of the class include, from the stomach, pepsins A and C and chymosin; from the lysosome, cathepsins D and E; from the kidney, renin; from yeast granules, proteinase A; from fungi, rhizopuspepsin, penicillopepsin, and endothiapepsin; and the retroviral proteinases from the HIV-1 and Rous sarcoma virus.[1,2] The retroviral proteinases, whose size is less than one-half the size of pepsin, function as dimers.[2] Since zymogens are known or have been inferred from gene sequence for all except the latter two fungal enzymes,[3] it is reasonable to assume that all the aspartic proteinases are synthesized as precursors. Except for the retroviral proteinases which are synthesized as part of the long *gag–pol* polypeptide, each precursor has an N-terminal extension of about 45 residues which is excised upon activation. One possible control mechanism for the expression of aspartic proteinase activity is the activation of the zymogen. In the case of pepsinogen, the precursor is synthesized and stored in the chief cells of the stomach wall. Activation does not occur until the pepsinogen is secreted and encounters hydrochloric acid which is secreted from parietal cells of the stomach.[4] Consequently, pepsinogen neither digests itself nor does harm to the host tissue while it is stored waiting for the stimulus of a meal. The storage and activation conditions of the other aspartic proteinase zymogens are not so well understood as that of pepsinogen. However, the zymogens of the other stomach enzymes, cathepsin D, renin, and rhizopuspepsin, are all able to activate with only a lowering of pH.[3–5] The HIV-1 proteinase is thought to excise itself from a long polypeptide before cutting that same polypeptide to form three separate proteins. One purpose of our study of the pepsinogen crystal structure was that it serves as a model for other aspartic proteinase zymogens. The activation rate constants for rhizopuspepsinogen are almost identical to those for pepsinogen.[6] Consequently, pepsinogen certainly is an excellent model for this fungal aspartic proteinase precursor.

Porcine pepsinogen contains a single amino acid chain of 370 residues. Upon activation, 44 residues are removed from the N-terminal end of the pepsino-

Received April 25, 1991; revision accepted August 1, 1991.
Address reprint requests to Jean A. Hartsuck, Ph.D., Oklahoma Medical Research Foundation, 825 N.E. 13th Street, Oklahoma City, OK 73104.

gen molecule.* There are three disulfide bonds in pepsinogen, one between residues 249 and 282, and two small loops formed by bonds between residues 45 and 50 and between residues 206 and 210. Pepsinogen activates spontaneously when exposed to pH less than 4. This removal of the activation peptide may be catalyzed either by a previously formed pepsin molecule (a bimolecular or autocatalytic mechanism) or intramolecularly. The intramolecular activation is apparently catalyzed by the nascent pepsin active site. Chemical identification of products has demonstrated that unimolecular cleavage occurs between residues 16P and 17P, with subsequent bimolecular removal of residues 17P through 44P.[7] There are also experiments which support the suggestion of unimolecular cleavage of the 44P-1 bond so that an intact activation peptide is released.[8] Another aim of this study is therefore to understand the mechanisms of pepsinogen activation.

Three coordinate sets of porcine pepsin from independent refinements are currently available.[9] The coordinates of pepsin crystallized from ammonium sulfate at pH 3.6 were determined by Cooper et al.[10] The structure from alcohol crystallization at pH 2 was refined independently by Abad-Zapatero et al.[11] and by Andreeva et al.[12] In this work, for the most part, the pepsinogen crystal structure is compared to the pepsin structure determined by Cooper et al.[10] The rationale for this choice is that the salt and pH of the crystallization solution are closest to that of pepsinogen. Consequently, differences between the structures are more likely to be real and not the result of different crystal environments. Crystal structures of the three above mentioned fungal aspartic proteinases have been known for some time.[13–15] Also available is a high-resolution structure of recombinant chymosin.[16] All of these structures are very similar to one another and demonstrate that the molecules are bilobal with a 2-fold symmetry between the N- and C-terminal lobes. This symmetry is also evident in the amino acid sequences of the two lobes and is especially apparent in the sequences around the two active site aspartic acid residues. The sequence Asp-Thr-Gly occurs at both active site aspartic acid residues of all known aspartic proteinases and is even present in the more distantly related retroviral proteinases which have an amino acid chain corresponding to one lobe of the other members of the class but function as a dimer. A preliminary report of the determination of the pepsinogen crystal structure has already appeared.[17] Our results confirm and considerably expand on the observations in that report. The present highly refined model of pepsinogen has been used as the basis for the solution of the structure of two other mammalian aspartic proteases, pepsin[10] and chymosin.[16] It was hoped that comparison of the pepsinogen and pepsin crystal structures would suggest causes for the difference in pH stability of the two molecules since pepsin is denatured at neutral pH, while pepsinogen is stable.

*Pepsinogen is numbered 1P through 44P for the activation peptide and 1 through 326 for the pepsin part of the molecule. The original determination of the amino acid sequence of pepsin placed an Ile residue at position 230 of pepsin. This residue does not appear in a subsequent gene sequence determination and is not present in our crystal structure. Consequently, in this report, all residues after 230 have had one subtracted from their number so that the amino acid sequence has continuous numbering. Also, the three known fungal proteinase crystal structures have been superimposed by rotation and translation onto the pepsinogen structure. Then the sequence of each fungal pepsin was renumbered in pepsin numbering. Of course, some insertions and deletions were required in these numbering schemes.

METHODS

Crystallization and Data Collection

Porcine pepsinogen, purchased from Sigma Chemical Company, was crystallized by either batch or vapor diffusion methods at 4° with final concentrations of 2.95% protein, 52% saturated lithium sulfate, and 0.025 M sodium cacodylate, pH 6.1.[18] The crystals were stored in a stabilizing solution containing 65% saturated lithium sulfate, 0.025 M sodium cacodylate, pH 6.0. The crystals belong to space group $C2$ with unit cell dimensions $a = 106.1$, $b = 43.7$, $c = 88.9$ Å, and $\beta = 91.4°$. There is only one molecule in the asymmetric unit of this crystal form. The chloroplatinate derivative was prepared by soaking pepsinogen crystals in stabilizing solution which contained 1 mM K_2PtCl_4.

Statistics for the original native and the chloroplatinate derivative data sets are included in Table I. The data were collected by means of oscillation photography on an Arndt-Wonacott camera and, after digitization on an Optronics P-1000 scanner, were reduced with the ROCKS Crystallographic Computing programs.[19] The chloroplatinate derivative data set was scaled to the native data using Wilson statistics and local scaling.[20] In the 20 to 3 Å resolution difference Patterson map, the four highest peaks corresponded to a two atom solution of the map. Centric refinement of this derivative yielded an R-factor of 0.43.

Late in the refinement stage of this structure determination, it was decided to recollect the native data. Data were collected using oscillation photography on a Rigaku RU-100 operated at 150 mA, 40 kV with a focal spot of 0.5×5.0 mm. The beam was filtered with a graphite monochromator. Crystals were mounted along a^* or c^* and oscillated through a range which varied from 2–3 degrees depending upon orientation. Cylindrical cassettes with a radius of curvature of 80 mm allowed data to be collected to 1.6 Å resolution perpendicular to the rotation axis, but the crystals were somewhat smaller than desirable so reflections beyond 1.65 Å had to be dis-

TABLE I. Data Collection Statistics

Compound	Resolution (Å)	Crystals	Films	R_{merge}	Reflections		Completeness		Resolution limits (Å)
					Total	Unique	Overall (%)	Shell (%)	
Native	2.55	6	131	0.112	36642	13083	95.5	93.6	3.00–2.55
$PtCl_4^{2-}$*	3.00	5	55	0.134	N.A.†	7665	88.2	N.A.	
Native	1.65	7	42	0.082	68120	31236	62.2	88.8	30–3.30
								89.4	3.30–2.62
								82.4	2.62–2.29
								71.4	2.29–2.08
								60.8	2.08–1.93
								50.6	1.93–1.82
								35.0	1.82–1.73
								18.4	1.73–1.65

*For the two $PtCl_4^{2-}$ binding sites the fractional coordinates are as follows:

x	y	z	Relative occupancy
0.7587	0.4230	0.3528	1.00
0.6804	0.4900	0.2145	0.65

†N.A., not available.

carded. Films were processed after digitization on an Optronics P-1000 optical scanner using the software described by Schmid et al.[21] Statistics for this data set are given in Table I.

Molecular Replacement Solution

Rotation function calculations were performed with the program written by Crowther.[22] Compared to including only lower resolution data, better discrimination was achieved when data from 10 to 4 Å was included in the calculations. Three models were employed for rotation function searches of the pepsinogen data. First was the entire pepsin molecule from an intermediate stage of refinement to 2 Å as provided by Dr. Natalia Andreeva, Institute of Molecular Biology, Academy of Sciences of the USSR, Moscow. Second was the same pepsin model with 43 residues deleted (SMPEP). Residues were deleted if their α-carbon atom was not within 5 Å of the homologous penicillopepsin α-carbon atom after superposition of the two models. Third was a penicillopepsin model formed by including atoms from the penicillopepsin structure[55] which were predicted to be the same as the corresponding atom in pepsinogen. The electron ratios and map statistics, respectively, for the three models, are 0.74 and 1.43, 0.64 and 1.64, and 0.54 and 0.94, where electron ratio is the ratio of the number of electrons in the model to the number of pepsinogen electrons and where map statistic is the ratio of the height of the correct peak in the rotation function to the highest incorrect peak. Better results for the smaller but more accurate model (SMPEP) are consistent with the observation that a correct polyalanine structure is sufficient in some cases for successful molecular replacement.[23]

Translation analysis was performed by a combination of packing tests and a T1 translation function.[24] The highest peak in the translation function was in a position allowed by packing analysis. This peak was 1.26 times the next highest peak, which incidently was not allowed by packing analysis. A phase set based on the properly oriented and translated SMPEP model was calculated. This set was used to phase an unweighted, locally scaled, 20–4 Å resolution, difference Fourier map for the chloroplatinate heavy atom derivative. The two highest peaks in that map had coordinates agreeing to within 0.3 Å with the previous Patterson map solution for the derivative. The pepsinogen electron density map based on the SMPEP phase set had well-defined molecular boundaries but was not readily interpretable. Consequently, a method for improving the phasing was devised and is described below.

Multiple Molecular Replacement

Interpretable pepsinogen electron density maps were calculated using a new phasing procedure. Since there were several models available from which to calculate phases, it seemed reasonable to combine model phases from the different models with heavy atom phases in a manner similar to that proposed by Hendrickson and Lattman[25] for combining phase information from various sources. A program was written to combine heavy atom phases with phases from up to three different atomic models. The phase probability distributions from the various sources were multiplied together as suggested by Hendrickson and Lattimann. Map coefficients were $2F_o–F_c$ in the usual manner, except F_c (and the phase of F_c, if a Sim-weighted map[26] was to be calculated) was derived from the three models by vector addition. It was found to be marginally better

TABLE IIA. Peak Heights in a Heavy Atom (HA) Difference Map for Rat Mast Cell Proteinase II (RMCP II)*

Model	*R*-factor (5–4 Å)	HA peaks (peak/rms)
GCHT	0.475	7.02, 6.20
ACHT	0.475	7.66, 6.20
EST	0.475	9.30, 7.13
GCHT + EST	0.442	9.33, 7.82
ACHT + EST	0.446	10.00, 7.62
GCHT + ACHT	0.462	7.90, 6.55
GCHT + ACHT + EST	0.438	9.44, 7.73
RMCP II	0.120	15.11, 14.73

*Phases were calculated from models of γ-chymotrypsin (GCHT), α-chymotrypsin (ACHT) and elastase (EST) by vectorial addition of model structure factors (see text).

TABLE IIB. Height of Major Heavy Atom Site in Pepsinogen Chloroplatinate-Native Difference Map.*

Model	*R*-factor (5–3.5 Å)	HA peak (peak/rms)
RHP	0.543	8.79
RHPR (refined)	0.466	8.98
PNP	0.542	4.97
PNPR (refined)	0.439	5.18
PGN initial	0.393	8.45
RHPR + PGN	0.445	10.05
RHPR + PGN SIM WT	—	10.22
PNPR + PGN	0.446	7.88
RHPR + PNPR	0.466	8.51
RHPR + PNPR + PGN	0.455	9.51
RHPR + PNPR + PGN SIM WT	—	9.63

*Phases were calculated using models of rhizopuspepsin (RHP), RHP refined for 9 cycles against pepsinogen data (RHPR), penicillopepsin (PNP), PNP refined for 9 cycles against pepsinogen data (PNPR), and the pepsinogen initial model (PGN). The summations indicate vectorial addition of the given structure factors. SIM WT indicates Sim-weighted vectorial addition (see text).

to perform weighted vector addition using Sim weights (see following discussion). The models used for phasing were the starting pepsin model (SMPEP),[14] highly refined coordinates of penicillopepsin[9] obtained from the Protein Data Bank,[9] and highly refined coordinates of rhizopuspepsin, obtained from D. Davies.[13]

This procedure, which we call "multiple molecular replacement," resulted in a greatly improved electron density map. Table II illustrates the efficacy of this procedure for a test case, rat mast cell proteinase II, a chymotrypsin-like proteinase[27,28] and for this study. In Table IIA, the peaks in an $F_{ph}-F_p$ (heavy atom-native) difference map phased with model phases are given after vector addition of structure factors of various combinations of γ-chymotrypsin,[29] α-chymotrypsin,[30] and elastase[31] (coordinates obtained from the Protein Data Bank and placed in the same orientation as RMCP II using the OVRLAP program of W.S. Bennett). It is clear that the sum of three models results in higher peaks than any of the single models, but that there is a best pair of models which results in the highest heavy atom peaks. In Table IIB, similar results are obtained for the unknown pepsinogen crystal structure. Here, all models were refined to convergence at 3.5 Å resolution against the pepsinogen data using the TNT program.[32] We concluded that the best combination of models was the initial pepsinogen model and the rhizopuspepsin model, where each had been subjected to nine cycles of coordinate refinement using TNT. Thus, for the initial electron density map, there were three sources of phase information, a heavy atom derivative and two protein models.

Model Building and Refinement

In early stages of the refinement, it was found to be useful to simultaneously inspect both combined-phase maps and Sim-weighted maps. Apparently, the different sorts of errors in the two phase sets affected different parts of the map. In some cases, side chain density would be clear in one map but not in another. In intermediate stages of the refinement, $2F_o-F_c$ maps were calculated from the pepsinogen model using Sim weights; in the final three cycles, no weighting was performed and both $2F_o-F_c$ and F_o-F_c maps were inspected.

Model building was performed using FRODO to construct and modify molecular models[33] on an Evans and Sutherland PS330. As an aid to map interpretation, the three available models of the fungal acid proteinases were superimposed on the initial pepsinogen model using the program OVRLAP. These models were then placed in the background (in different colors), allowing simultaneous comparison of the electron density and four models. It was often the case that one of the fungal acid proteinases fit a particular segment of the pepsinogen electron density map much better than the other two (or than the initial pepsinogen model) due to insertions or deletions. This technique proved to be very helpful in suggesting alterations to the initial pepsinogen model. Water molecules were added in the final stages where there were peaks larger than 3.5 standard deviations in the F_o-F_c map and which were in appropriate locations to make reasonable hydrogen bonds to other atoms.

Automated crystallographic refinement was performed using the TNT program package[32] and at one point EREF.[34] Structure factors were calculated using the program GENERALSF by Tronrud, which automatically takes into account as many of the crystallographic symmetry operators as possible, using operator-specific fast Fourier transform (FFT) subroutines. Both of the refinement procedures es-

TABLE III. Course of the Refinement of the Pepsinogen Structure*

Cycle	Active atoms	Resolution (Å)		Final *R*-factor	Total shift (Å)
		Refine	Maps inspected		
	2236	10–2.6		53.2	0
1	2236	10–3.0		43.5	0.3
2	2553	5–3.5	4.0 PHC	31.9	2.76
3	2558	5–3.0	3.5 PHC/5–3.5 Sim	34.4	2.91
4	2629	5–2.55	3.5 PHC/5–3.5 Sim	29.5	3.17
5	2636	5–2.55	3.0 PHC/5–3.0 Sim	27.3	3.09
6	2599	5–2.8	3.5 PHC/5–2.8 Sim	28.2	—†
7	2618	5–2.55	5–2.8 Sim	27.4	—
8	2654	5–2.55	15–2.8 Sim/solvent	22.9	3.14
9	2684	5–2.55	25–2.55 Sim/solvent	22.3	3.23
10	2723	5–2.55	25–2.55 Sim/solvent	21.5	3.45‡
11	2789	5–2.55	25–2.55 Sim/solvent	20.5	
12	2799	5–1.8	25–1.8 Sim/solvent	24.3	
13	2828	5–1.7	5–1.7 Sim	22.8	
14	2832	5–1.7	5–1.7 $2F_o-F_c/F_o-F_c$	20.5	
15	2849	5–1.65	5–1.65 $2F_o-F_c/F_o-F_c$	18.6	
16	2872	5–1.65	5–1.65 $2F_o-F_c/F_o-F_c$	17.3§	

*Total shift is the root mean square shift of all atoms with identical names and residue numbers to the starting model. Solvent indicates that structure factors calculated from the solvent region were vectorially added to the low-resolution protein structure factors. PHC indicates phase combination of single isomorphous with calculated phases, SIM indicates that Sim-weighted density maps were inspected.

†Models lost due to malfunction of magnetic tape unit.

‡Eleven cycles EREF refinement.

§Data where $F_{obs} < 1.3\ \sigma\ (F_{obs})$ were discarded. Models 11–16 all have shifts of approximately 3.5 Å from the starting model. They differ chiefly in the solvent structure and orientation of some side chains.

timate crystallographic gradients from an FFT difference electron density map and are thus very fast. The philosophy followed in this investigation was to allow the model to become quite distorted in order to allow large movements, then a number of cycles were run with increased weight on ideal geometry until the final stereochemistry was acceptable.

Sixteen macrocycles of manual model-building and automated refinement were required to obtain the final model, with each round of automated refinement requiring between 10 and 50 cycles of TNT (see following paragraph). The course of the pepsinogen refinement is summarized in Table III. In later stages, temperature factors were refined for 3–5 cycles after 3–5 coordinate shift cycles. Solvent molecules were discarded in intermediate stages if the temperature factor rose above 50 Å^2.

Between macrocycles 5 and 10 it was noted that the convergence of the TNT program was very slow and the results (*R*-factor versus stereochemistry) were unsatisfactory. For cycle 11, the Jack–Levitt EREF procedure was used and showed very rapid convergence. After some effort, it was discovered by Dr. Tronrud that the spacial derivatives were inaccurate due to improper gridding of the difference density maps. Although a grid spacing of one-third of the resolution had been chosen, this was too coarse. Apparently, at least in the 2.5–3.0 Å resolution range, a grid spacing of about one-quarter of the resolution must be used. Dr. Tronrud now recommends a "blurring factor" of at least 20 for coordinate as well as temperature factor refinement.[32]

A dramatic improvement in the appearance of the electron density maps was achieved after cycle 7 by including structure factors of the bulk solvent region in the low resolution data (25–5 Å). The procedure used is that discussed by Remington et al.[35] The quality of the $2F_o-F_c$ pepsinogen electron density map can be appreciated by examination of Figure 1. Figure 1A shows the region of the molecule surrounding the enzyme active site, while Figure 1B demonstrates the density for alternate conformations of the side chain of Thr-17. The final pepsinogen model has been deposited in the Protein Data Bank[9] as entry 1PSG. The coordinates include 365 amino acids (excluding amino acids 157–161) and 180 solvent molecules.

RESULTS

Completeness and Accuracy of the Model

The final model of pepsinogen has an *R*-factor of 17.3% for the 28,806 reflections of a magnitude greater than 1.3 σ and resolution between 5 and 1.65 Å. Stereochemistry is excellent with rms deviation from ideal bonds of 0.015 Å and from ideal bond angles of 2.7°. The model contains 180 solvent molecules. There is no electron density for residues 157–161; these extend into the solvent and are apparently disordered. The final F_o-F_c map has a large positive peak of about 0.8 electrons/Å^3 near the α-

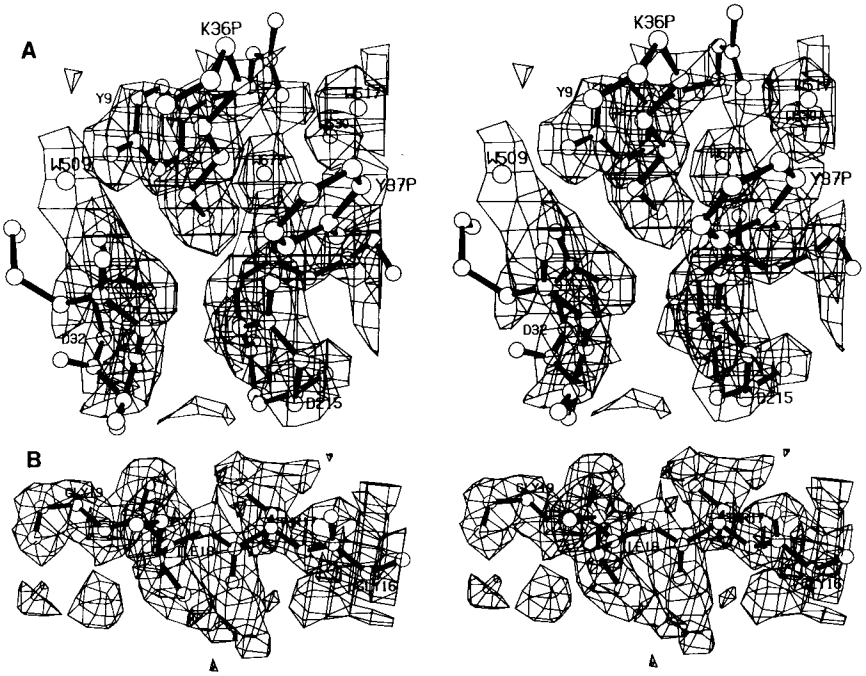

Fig. 1. Two stereo views of the $2F_o-F_c$ electron density map of pepsinogen contoured at 1 standard deviation. (**A**) In the vicinity of the catalytic aspartyl residues and Lys-36P. The final refined model was used for calculation of phases to 1.7 Å resolution. (**B**) In the vicinity of Thr-17. The map shows clear evidence for multiple conformations of Thr-17.

carbon of residue 162. This peak is considerably larger than a glycine residue dictated by the amino acid sequence.[36] It has not proved possible to interpret this density with reasonable stereochemistry and in the present model was left as glycine. The rest of the polypeptide backbone is in interpretable density, but several areas are weak and of doubtful accuracy. Regions of the backbone where density is weak are 42P-1, 46–49, 76–78 "flap region," 128–131, 157–161, 235–244, 252–256, and 290–297. Interestingly, all of these are loop regions on the surface of the protein and in some cases extend quite far away from the rest of the surface. Most dramatic are the regions 42P-1 where the final cleavage occurs in the activation step, 235–244 and 290–297. Indeed, residues 290–297 are supported by a tenuous intermolecular contact and are probably free to take on other conformations when the molecule is in solution. Table IV lists side chains for which there is either weak or uninterpretable density. Notably, the phosphate group of phosphoserine 68 is not visible in the electron density maps. A Luzzati plot[37] is shown in Figure 2, for the R-factor calculated in 10 shells of resolution from 5.0 to 1.65 Å. From this, it appears that the rms coordinate error of this investigation is about 0.2 Å, although the nonlinearity of the plotted points indicates this is a crude estimate. Figure 3 shows a ϕ,ψ diagram[38] for pepsinogen. There is a single nonglycine residue, Thr-293, far outside the acceptable region. This residue is in the weakest density of the loop 290–297, with no density beyond the β-carbon; thus its conformation is almost certainly in error.

For atoms included in the model, the mean thermal factor for 1460 main chain atoms is 26, and for 1,232 side chain atoms is 29. In Figure 4, temperature factors, averaged over the main chain atoms, are plotted as a function of sequence number. Where the thermal factor exceeds 40, the residue names are indicated in the plot. It is very interesting that four of the peaks occur adjacent to disulfide bridges. Not only were the thermal parameters elevated in the vicinity of disulfide bridges but also the electron density was visibly weak in these regions. Ser-46 is next to the 45–50 disulfide, the loop at Asp-200 is near the 206–210 disulfide, Ser-254 and 278–281 are next to the 249–282 disulfide. Most of the propeptide seems to have large thermal factors and is probably weakly bound to the proteinase. The loops 239–244 and 292–295 extend away from the molecule, making few other stabilizing interactions. The regions of highest disorder in pepsinogen correspond to those of highest mobility in pepsin, 234–256, 278–283, and 294–298. The "flap regions" of pepsinogen, pepsin, and penicillopepsin have high temperature factors, but not as high as other zones of the molecule. This is in contrast to the rhizopus-pepsin structure, where the flap is the most mobile region.[13] While such high temperature factors can be the result of grossly incorrect interpretation of

TABLE IV. Pepsinogen Coordinates of Doubtful Accuracy*

Residue	Atoms	Comments
3P	CE,NZ	
9P	CD,CE,NZ	
10P	NZ	Side chain may take multiple conformations
14P	NE2,OE1	
21P	CB,CG,CD,CE,NZ	
23P	CD,CE,NZ	
24P	CB,(CG,OD1,OD2)	
27P	CG,CD,CE,NZ	
30P	CD,CE,NZ	
39P-1		Very weak density for main chain only
39P	CA	
40P	C,CG,CD,OE1,OE2	
42P	CB	
43P	CB	
44P	CB,CG,CD1,CD2	
1	CB,CG2,CG1,CD1	
17		Side chain has two conformers†
46–47		Weak but continuous main chain density
51		Weak but continuous main chain density
68		Phosphoserine, very weak density beyond CB
70	CG,CD,OE1,OE2	
73	CB,CG2,CG1,CD1	
75		Weak main chain density
81		Weak main and side chain density
83		Side chain coordinates are corrected‡
90		Weak side chain density
132–133		Weak main chain density
143		Interchange atom labels OE1–NE2
157–161		No interpretable density§
176	OG1,CG2	
178	OG	
180	ND1,OD1	
186–187		Weak but continuous main and side chain density
199–202		Weak main chain density
199	C	
200	OD1,OD2	
208–210		Weak main chain density
225	CG2,(OG1)	
226	OG	
227	CD1	
238	CB,OG	
239	CG,CD,OE1,OE2	
240–243		Weak main and side chain density
241	OG	
242	N,CA,C,O,CB,CG,OD1,OD2	
243	N,CA	
253	OD1,OD2	
254	CA,(CB,OG)	
263	CB,(CG,OD1,OD2)	
266	CG,CD,(OE1,NE2)	
277	OE1,(NE2)	
278	CB,CG,OD1,OD2	
279	N,CA,C,O,CB,CG,OD1,OD2	
280	N,CB,CG,OD1,OD2	
281	OG	
292–296		Weak main chain density
292	N,CA,CB,CG,CD	
293	O,OG1	
294	CB	
295	CB,OG	

*The final $2F_o-F_c$ electron density map was contoured at a cutoff of one standard deviation (0.15 e/Å^3), and atoms in regions of density lower than this are listed in the table. Their positions should be considered hypothetical and are based on other well-localized atoms which are nearby. Atoms in parentheses are in density higher than this cutoff, but may have incorrect coordinates for other reasons. Comments flag regions where the electron density is only slightly higher than the cutoff, and may also be considered weak. In some cases, the electron density was so weak that the side chain was replaced with a smaller side chain in the model.

†Alternative coordinates for Thr-17 OG1 (55.75,−14.22,20.85), CG2 (58.02,−13.68,20.48). We estimate each conformer to have 50% occupancy based on the electron density map.

‡The deposited side chain coordinates of Ile-81 are in error. The corrected coordinates (in the same coordinate system as the deposited atomic model) are CB (66.60,1.45,34.37), CG2 (66.93,18.5,35.82), CG1 (66.59,−0.08, 34.35), CD1, (67.76,−0.67,33.56).

§An alternative interpretation for residues 162 and 163 is preferred over that of the deposited coordinates. In the deposited coordinates residue 162 occupies density which we now interpret to be the side chain of 163. The preferred coordinates are N 162(45.95,−16.45,9.62), CA 162 (46.34,−15.88,10.91), C 162 (46.29,−14.35,10.84), O 162 (46.16,−13.76,9.75), N 163 (46.36,−13.76,12.01), CB 163 (45.66,−11.93,13.47), OG 163 (44.54, −12.78,13.68).

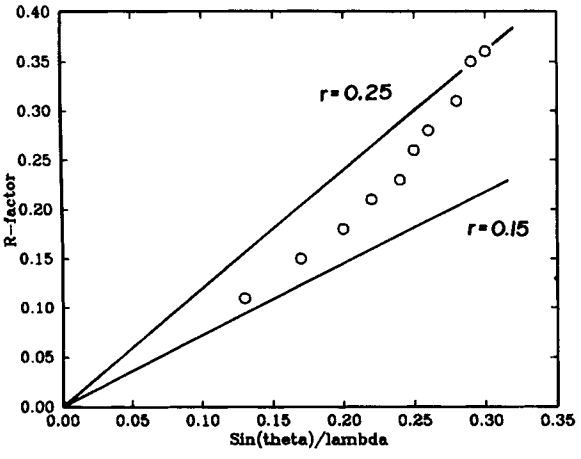

Fig. 2. Luzzati plot of *R*-factor versus sin(θ)/λ. The solid lines indicate theoretical variation of *R*-factor for rms coordinate errors of 0.15 and 0.25 Å.

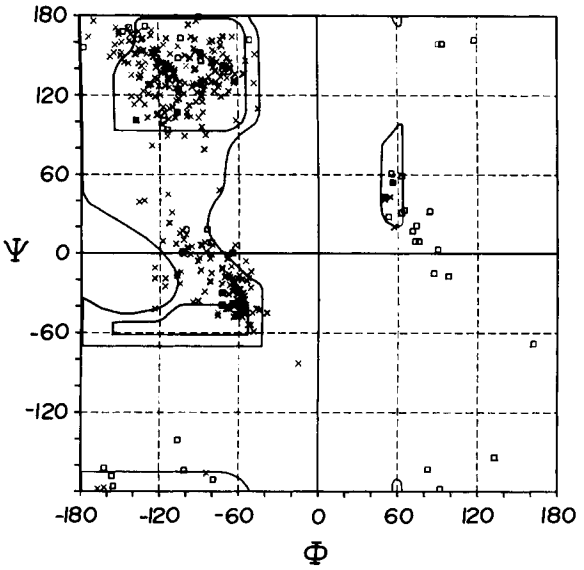

Fig. 3. Ramachandran diagram of the main chain dihedral angles of the refined atomic model of pepsinogen.

the electron density, we do not think this is the case here. Electron density maps, with these residues removed for the purpose of the phase calculation, do show continuous but weak density at all of these locations.

Conformation of the Molecule

Pepsinogen has a rather pronounced ellipsoidal shape with axial lengths of about $30 \times 50 \times 65$ Å. As was expected, the pepsin part of pepsinogen strongly resembles that of pepsin and those fungal acid proteinases whose structures are known, consisting largely of β-sheet. α-Carbon traces of the backbone are shown in Figure 5A and B. Excluding the propeptide, the molecule has an approximate 2-fold axis of symmetry which suggests that the molecule evolved from a dimeric structure via a gene-duplication event.[39] The walls of the active site cavity, which is cavernous, are formed from two compact domains, each of which shows additional 2-fold symmetry as has been previously discussed.[40,41] Using this OVRLAP program, of the roughly 130 residues in each domain, 63 α-carbon positions superimpose with an rms error of 2.8 Å. Figure 6 shows this superposition. It is interesting to note that the "flap region," residues 70 to 85, which is thought to be catalytically important, is an insertion into the basic topology of either domain. In Figure 6, the "flap," a β ribbon which lies over the substrate binding cleft and is common to all aspartic proteinases, extends prominently out toward the viewer in the center of the drawing to extend toward the active site.

The N- and C-terminal domains make relatively few contacts with one another and are supported "under" the active site by a large six-stranded β-sheet, consisting of the N- and C-terminal strands from each of the two domains. This β-sheet obeys the same approximate two-fold symmetry as do the two domains. Figure 7 is a stereo view of the backbone and β-carbon atoms of this sheet with the point of view approximately along the 2-fold axis. The covalent connection between the N- and C-terminal domains is the strand 169–173 which crosses behind the sheet on the outside of the molecule.

The propetide residues, 1P–44P, comprise three helices and one strand of a β-sheet, with short irregular peptides connecting the secondary structure elements. The N-terminal residues of pepsinogen (2P–9P) form the outer strand of the β-sheet shown in Figure 7. This position is occupied by the N-terminal residues of the active proteinase structures.[10,11,13–15] The helices pack into the active site crevice in a manner which does not completely fill the crevice. Most of the interactions of the propeptide are with the N-terminal wall of the active site cavity (Fig. 5B).

Secondary Structure in the Molecule

Within the pepsinogen molecule, regions of secondary structure have been demarcated by requiring the existence of the characteristic hydrogen bonding pattern. The criteria used to define hydrogen bonds are those established by Baker and Hubbard.[42] The total quantity of regular secondary structure (β-structures, α-helices, and 3_{10}-helices) in the activation peptide and pepsin parts of pepsinogen are very similar (68 and 71%, respectively). However, helical structures are much more prevalent in the activation peptide (75% of secondary structure therein) whereas extended chains predominate in the pepsin part of the molecule (69% of secondary structure therein). In retrospect this is probably not surprising since the activation peptide could be more easily excised if it contained consec-

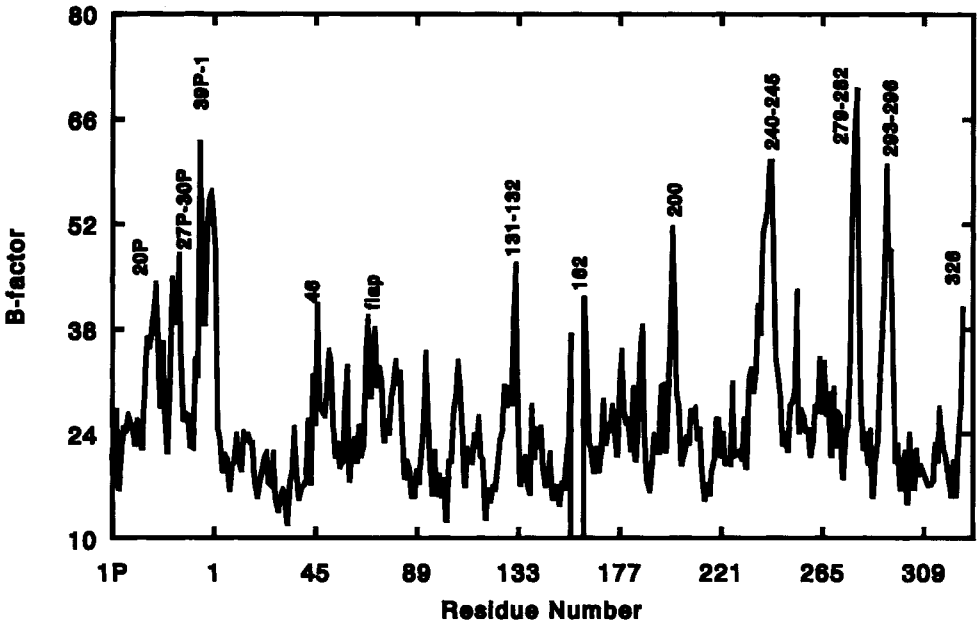

Fig. 4. Thermal factors (Å^2) averaged over the main chain atoms in each residue plotted as a function of the residue number. Regions of high mobility are identified by their amino acid sequence numbers. Residues 157–161 are disordered.

utive interactions (i.e., helix) rather than interactions with a more remote portion of the primary structure.

The α-helices and β-structures of pepsinogen are represented in Figure 8. The pepsinogen molecule consists largely of β-structure with six α-helices. The 2-fold relationship which exists within the pepsin portion of the molecule is apparent in this diagram. The strands of β-structure in the two lobes are related to one another as can be deduced from Figures 6 and 8. However, there is no element present in the opposite lobe to match residues 65–74 and 80–87 of the N-terminal lobe (see also Fig. 6) or 238–241, 243–247, 272–277, and 280–284 of the C-terminal lobe. The degree of perfection of the various β-sheets can be appreciated by examining the hydrogen bonding patterns documented in Figure 8. The helices from 135 to 143 and from 302 to 309 are symmetrically related to one another, and, in fact, portions of those helices overlap when the N- and C-terminal domains are aligned. There are two other α-helices in the C-terminal lobe, residues 224 to 235 and 270 to 275, which do not have counterparts in the N-terminal lobe. Of course, the helices of the activation peptide, viz. 10P to 19P and 20P to 29P, do not have symmetrically related counterparts since the activation peptide does not obey the 2-fold symmetry of the rest of the molecule. Table V describes each of the pepsinogen α-helices and lists the average values of the parameters of the hydrogen bonds and dihedral angles. As shown in Table V, the mean values of ϕ and ψ for the pepsinogen α-helices are $-60°$ and $-41°$. A thorough study of helix geometry in 57 proteins by Barlow and Thornton[43] found mean values of $-62°$ and $-41°$. The longest helix is formed by the 12 residues 224 to 235 and has eight consecutive hydrogen bonds. Six stretches of the pepsinogen chain have been classified as 3_{10} helices. The location and properties of these helices are shown in Table VI. Mean values of ϕ and ψ for the pepsinogen 3_{10} helices are quite near the ideal values. Multiple turn 3_{10} helices are more uncommon than long runs of α-helix. In pepsinogen, the longest run of 3_{10} helix includes residues 109 through 115 with four consecutive hydrogen bonds. In the helix of residues 248 through 255, the four hydrogen bonds are not consecutive but rather are two sets of two hydrogen bonds. The short length and high incidence of 3_{10} helices in pepsinogen are consistent with the observation of Barlow and Thornton[43] that 3_{10} helices are in general short and more likely to occur in all-β proteins than in α/β proteins.

The reverse turns of pepsinogen are tabulated in Table VII. Classification of a turn as a particular type is not very stringent and demands only that a hydrogen bond be present between the first and fourth residues of the turn and that no more than two of the four angles ϕ_2, ψ_2, ϕ_3, and ψ_3 deviate from the ideal value by more than 40°.[44] Even so, the actual disagreement between observed and ideal ϕ and ψ angles is quite small. The reverse turns are located in almost all instances so that either the first or the fourth residue of the turn takes part in helix or β-structure. Only the turn involving residues 292

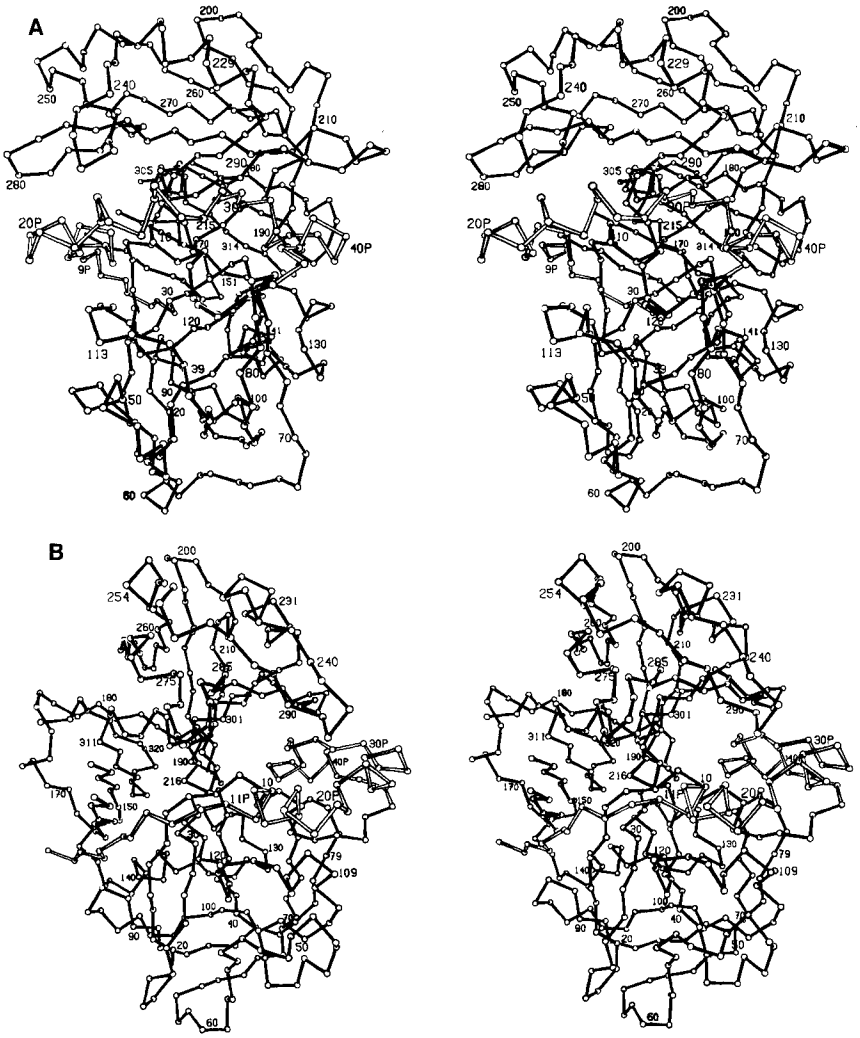

Fig. 5. Two stereo views of the α-carbon backbone of pepsinogen. The activation peptide is represented with open bonds. The second view was generated by approximately a 90° rotation about the vertical axis of the drawing. (**A**) The β-sheet which involves residues 2P–9P is on the left of the drawing. (**B**) The helices of the activation peptide, residues 10P–19P and 20P–29P, can be seen filling the substrate binding cleft of the nascent enzyme.

through 295 is located in a random coil segment of the chain. There are two small disulfide loops in pepsinogen between residues 45 and 50 and between 206 and 210. Within each of these loops, there is a reverse turn. Residue Pro-23 is in the *cis* conformation, and the backbone makes a turn at this point. The dihedral angles are consistent with the *cis*-proline turn of type VIb, however it was not included in Table VII because the relationship between backbone atoms of residues 21 and 24 did not meet the criteria for a hydrogen bond.

At the C-terminal end of α-helices a π turn hydrogen bonding pattern can occur. In this case, there is a reverse turn within a loop defined by a hydrogen bond from the carbonyl oxygen of residue n to the imide of residue $n + 5$. Three such conformations occur in pepsinogen. In two of the three, the interior reverse turn is of type I, and, in the third instance, it had to be included in the type IV (miscellaneous) category.

Comparison of the structure of pepsinogen to pepsin shows surprisingly few changes in secondary structure (outside of those regions where activation results in an overall conformational change, viz. 1P–13). In regions of well-defined secondary structure, there exist subtle changes in backbone dihedral angles between the two structures which result in the definition of a hydrogen bond in one structure, but not the other. The lack of a hydrogen bond in those cases is not evidence for a gross change in secondary structure and might be due to barely meeting or failing the hydrogen bond criteria. However, one region, 290–297, shows a drastic difference in conformation. This region in pepsin is a β ribbon. In pep-

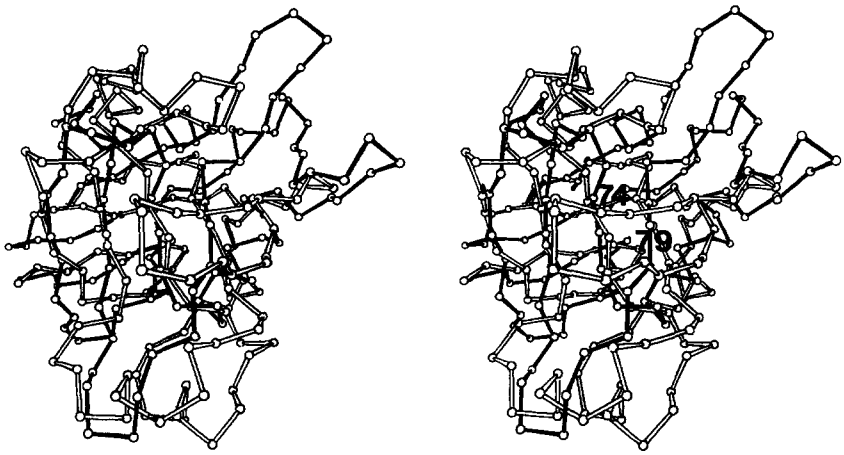

Fig. 6. Stereo view of the optimal superposition of the N-terminal (open bonds) and the C-terminal (closed bonds) domains of pepsinogen. Note the close alignment of the chains in the core of the two domains. Insertions or deletions are on the surface of the domains. The "flap region" extends toward the viewer and is identified by labels on α-carbon atoms 74 and 79.

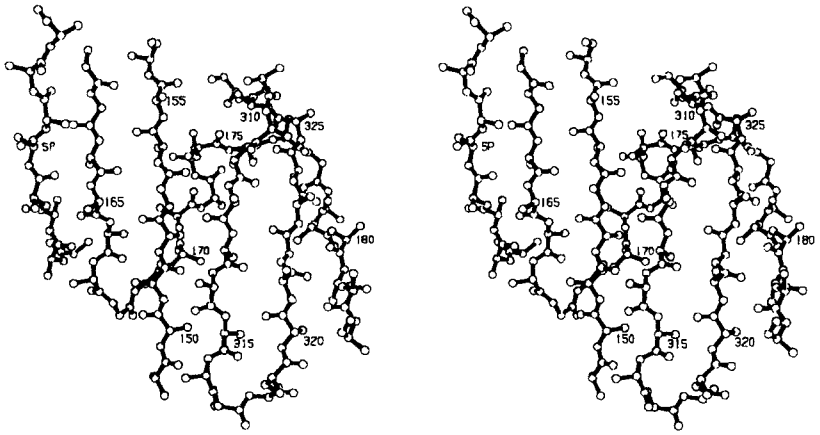

Fig. 7. Stereo view of the β-sheet which connects the N- and C-terminal domains. Side chain atoms beyond β-carbon atoms have been removed. The drawing demonstrates the local 2-fold symmetry characteristic of the N- and C-terminal domains. The break between residues 156 and 162 is due to a disordered segment of the chain.

sinogen, these residues form a loose loop with no cross-chain hydrogen bonding. However, pepsinogen does have hydrogen bonds from this loop to external chains. This region interacts with the activation peptide, in the vicinity of the P1′ binding subsite, and exhibits high temperature factors.

Hydrogen Bonds Involving Side Chains

Not including the ion pairs, which are discussed separately, pepsinogen has 44 side chain to side chain hydrogen bonds and 67 side chain to backbone hydrogen bonds. These numbers do not include interactions with the solvent. This is compared to the 173 backbone to backbone hydrogen bonds. The pepsin structure surprisingly shows few differences in side chain hydrogen bonding. As was observed in the comparison of backbone–backbone hydrogen bonds, subtle changes in side chain dihedral angles, which result in a lack of definition of a hydrogen bond between a pair of residues, account for most of the differences. Outside of regions which interact directly with residues 1P–13 (which undergo excision or gross movement upon activation), no major differences in the sidechain hydrogen bonding network are observed. Glu-107 is an exception to this observation, as the change in hydrogen bonding is neither subtle nor due to interaction with residues 1P–13, but, in pepsinogen, is involved in an intermolecular contact (see discussion). The percentage of acidic side chains in pepsinogen (30 aspartic and 13 glutamic acids out of 370 residues) is greater than that for many proteins, and several of these acidic side

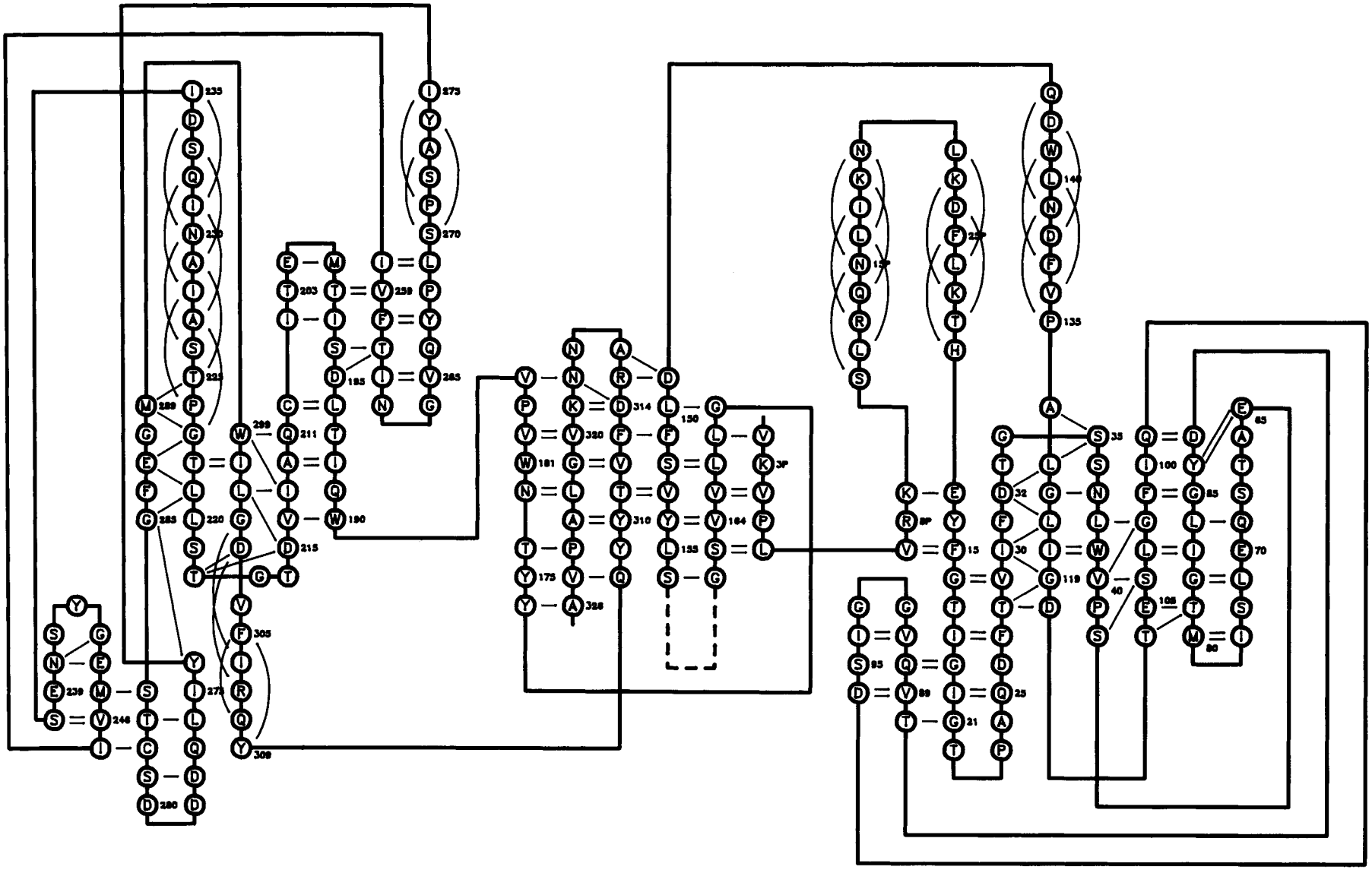

Fig. 8. Representation of the α-helices and β-structure of pepsinogen. Only residues involved in α-helices or β-structure are shown. Heavy lines represent the peptide chain connections and lighter lines represent hydrogen bonds. The dotted line shows the peptide chain (157–161) which was not seen in the pepsinogen electron density map. Residues 274, 275, 308, and 309 each appear twice in the diagram since they are involved in both α-helix and β-structure.

TABLE V. α-Helices

Residue	H-bonds	Average value for N..O (Å)	N..OC (°)	H..O (Å)	NH..O (°)	H..OC (°)	ϕ (°)	ψ (°)
11P–19P(9)	5	2.92	156	2.00	152	148	−62	−38
22P–29P(8)	4	3.05	162	2.12	154	155	−57	−46
135–143(9)	5	3.02	154	2.07	157	149	−60	−43
224–235(12)	8	2.97	157	2.01	160	151	−60	−42
270–275(6)	2	3.16	160	2.24	152	152	−65	−36
302–309(8)	3	3.06	152	2.14	154	145	−57	−38
Mean		3.03 (.08)*	157 (3)	2.10 (.08)	155 (3)	150 (3)	−60 (3)	−41 (3)
Ideal†		2.99		2.06	155	147	−60	−41

*Standard deviations are shown in parentheses.
†Ideal values are from Baker and Hubbard[42] and Barlow and Thornton.[43]

TABLE VI. 3_{10} Helices

Residue	H-bonds	Average value for N..O (Å)	N..OC (°)	H..O (Å)	NH..O (°)	H..OC (°)	ϕ (°)	ψ (°)
32P–37P(6)	3	3.05	126	2.10	161	122	−64	−20
6–10(5)	2	3.05	127	2.14	152	127	−57	−21
57–61(5)	2	3.19	112	2.24	158	106	−66	−24
109–115(7)	4	3.15	120	2.31	142	111	−61	−28
125–128(4)	2	3.04	124	2.06	164	123	−58	−29
248–255(8)	4	3.25	119	2.28	163	117	−63	−26
Mean		3.12 (.08)*	122 (5)	2.19 (.09)	156 (8)	118 (7)	−62 (3)	−25 (3)
Ideal†		3.09		2.17	153	114	−60	−30

* Standard deviations are shown in parentheses.
†Ideal values are from Baker and Hubbard[42] and Richardson.[44]

chains are involved in stabilizing interactions within the protein. As will be discussed in detail below, the two active site aspartic acid side chains are involved in a complex hydrogen bonding network. In addition, within the N-terminal lobe, Asp-87 binds to Thr-88 N (2.64 Å), to Ser-61 OG (2.63 Å), and Thr-63 OG (2.60 Å); Asp-118 binds to Asn-54 N (2.72 Å), Thr-28 N (3.21 Å), Asn-54 ND2 (2.38 Å) and is in an ion pair with His-53; and Asp-96 makes a hydrogen bond to Asn-139 ND2 (3.12 Å). Within the C-terminal lobe Asp-171 binds to Ser-173 N (2.89 Å); Asp-195 binds to Gly-264 N (2.99 Å); and Asp-314 makes three hydrogen bonds to Ala-316 N (2.64 Å), Asn-317 N (2.82 Å), and Asn-317 ND2 (1.95 Å).

As mentioned in the description of the structure above, the interactions between the two lobes of the molecule are quite sparse. Figure 8 shows five β-structure hydrogen bonds between the two lobes in the β-sheet at the center of the figure. The active site aspartic acids (32 and 215) and threonines (33 and 216) have several interlobe hydrogen bonds which are described below in the active site section. Besides these, there are only two hydrogen bonds and one ion pair interaction between the two lobes. Tyr-125 OH binds to Asn-318 OD1 (2.59 Å) and Ser-152 OG binds to Tyr-310 OH (2.62 Å). Reference to Figure 8 shows that residues 310 and 318 are both located in the sheet which is formed by strands from the N- and C-terminal residues of both lobes.

Ion Pairs and Buried Charges

The ion pairs of pepsinogen are particularly interesting because they may be important in the mechanism of activation of the molecule. Since intramolecular activation is triggered by a decrease in pH to a level which would discharge the usual side chain carboxyl group, destruction of stabilizing ion pairs may play a role of inducing the conformation change which accompanies activation. The ion pairs of pepsinogen and their properties are described in Table VIII. Eleven pairs of positive and negative side chains are close enough to one another to be designated ion pairs.[45] Nine of those interactions cannot exist in the active enzyme. However, inspection of the sequence alignment of the homologous aspartic proteinase precursors[46] shows that none of the ion pairs of Table VIII are conserved in all of the known sequences, with exception of the two interactions between Lys-36P and the two active site aspartates.

TABLE VII. Reverse Turns

	N..O (Å)	N..OC (°)	H..O (Å)	NH..O (°)	H..OC (°)	ϕ_2 (°)	ψ_2 (°)	ϕ_3 (°)	ψ_3 (°)	Location*
Type I										
38P–41P	2.85	133	1.89	162	128	−58	−5	−109	−18	C_{310}
4–7	3.05	132	2.07	163	135	−54	−24	−104	15	N_{310}
32–35	3.27	123	2.31	159	124	−70	−4	−101	14	C_β
50–53	3.31	130	2.37	156	132	−53	−41	−116	44	C_{310}
141–144	2.99	113	2.01	167	111	−57	−36	−88	8	C_α
233–236	3.31	117	2.38	155	110	−68	19	−97	−2	C_α
240–243	2.89	119	2.03	143	110	−54	−55	−63	4	Tight
Mean	3.10 (.2)†	124 (7)	2.15 (.2)	158 (7)	121 (10)	−59 (7)	−21 (24)	−97 (16)	9 (18)	
Ideal‡						−60	−30	−90	0	
Type I′										
91–94	2.99	119	2.03	159	119	63	36	84	−16	Tight
262–265	2.89	122	1.92	161	115	50	43	75	7	Tight
Mean	2.94 (.05)	120 (1)	1.98 (.05)	160 (1)	117 (2)	57 (6)	40 (4)	79 (4)	−4 (12)	
Ideal						60	30	90	0	
Type II										
129–132	3.01	124	2.15	143	130	−67	140	59	19	C_{310}
206–209	3.18	120	2.29	148	128	−68	121	74	22	SS§
Mean	3.10 (.08)	122 (2)	2.22 (.07)	146 (2)	129 (1)	−68 (1)	130 (10)	66 (8)	20 (2)	
Ideal						−60	120	90	0	
Type III										
16P–19P	3.11	112	2.23	147	102	−60	−49	−60	−18	C_α
42–45	3.09	114	2.13	159	108	−64	−23	−76	−23	C_β
47–50	3.08	122	2.13	157	115	−56	−37	−65	−20	C_{310},SS§
171–174	2.81	141	1.82	168	138	−52	−15	−73	−17	N_β
270–273	2.74	124	1.93	137	111	−58	−32	−63	−30	N_α
302–305	2.98	132	2.16	138	124	−47	−43	−46	−43	N_α
305–308	3.02	119	2.10	154	111	−61	−36	−57	−23	C_α
Mean	2.98 (.1)	123 (9)	2.07 (.1)	151 (11)	116 (11)	−57 (5)	−34 (11)	−63 (9)	−25 (9)	
Ideal						−60	−30	−60	−30	
Type III′										
130–133	3.10	126	2.10	175	125	59	18	53	28	C_{310}
Ideal						60	30	60	30	
Type IV										
17P–20P	2.68	118	1.75	154	112	−110	7	54	62	C_α
293–296	2.40	131	1.58	136	116	−16	−85	−70	2	R.C.
π turns										
16P–21P	2.87	142	1.98	148	133					C_α
140–145	2.87	157	1.88	169	160					C_α
232–237	2.95	161	1.95	171	164					C_α

*N_{310} and C_{310} are the N- and C-ends of 310 helix. N_α and C_α are the two ends of α- helix. N_β and C_β are the ends of β-strands. A tight turn is a reverse turn which connects two antiparallel β-strands. R.C. is random coil.
†Standard deviations are shown in parentheses.
‡Ideal values come from Richardson.[44]
§Residues 45 and 50 and residues 206 and 210 are joined by a disulfide bond.

Consequently, one cannot identify any additional particular interaction as the "key" to zymogen activation throughout the class of enzymes.

Accessibility calculations[47] demonstrate three buried aspartic acid residues (in addition to the active site aspartic pair). These three residues are not involved in ion pairs. Aspartic acid 87 makes three hydrogen bonds: OD1 to Ser-61 OG (2.63 Å), OD2 to Thr-63 OG1 (2.60 Å), and OD2 to backbone 88 HN (2.64 Å). These interactions are virtually identical in pepsin. The side chain of Asp-87 lies in a depression on the surface of the molecule and is not accessible to solvent in our static structure but could well have solvent contact in a dynamic structure. Aspartic acid 96 is hydrogen bonded to Asn-139 ND2 (3.12 Å). This residue is not well conserved in the aspartic proteinase family. Finally, Asp-303 OD1 is near (3.31 Å) Thr-218 O, but, even if the carboxyl group were protonated, a hydrogen bond would be unlikely because of a very distorted O–H..O angle.

TABLE VIII. Ion Pairs

Positive side chain	Accessibility (%)	Negative side chain	Accessibility (%)	Distance† (Å)	Not in active enzyme
Lys-3P NZ	33	Asp-171 OD2	18	3.29	*
Arg-8P NE,NH2	51	Glu-13 OE2	8	3.03,3.13‡	*
Arg-13P NE,NH2	12	Asp-11 OD2,OD1	13	2.80,2.79	*
His-29P NE2	20	Asp-3 OD1	32	2.93	*
His-31P ND1	18	Glu-7 OE1	7	2.81	*
His-31P NE2	18	Glu-287 OE1	20	3.82	*
Lys-36P NZ	7	Asp-32 OD1,OD2	0	3.52,3.17	*
Lys-36P NZ	7	Asp-215 OD1,OD2	0	3.46,2.81	*
His-53 ND1	15	Asp-118 OD1	6	2.90	
Arg-307 NH1	10	Glu-13 OE2	8	3.66	*
Arg-315 NH1,NH2	20	Asp-138 OD2,OD1	10	3.15,3.02	

*Side chain accessibility is calculated in the native structure with the program of Lee and Richards.[47] Side chains with accessibility less than 2% are considered buried.[14]

†The structure was searched for interactions less than 4.0 Å.[45]

‡When more than one interaction met the distance criteria, both atoms and distances are tabulated.

Interestingly, Asp-303 (which is well conserved in the gastric, fungal, and lysosomal aspartic proteinase sequences) is located near the active site and comparison with the pepsin structure shows a high degree of conservation of side chain orientation around Asp-303. In conclusion, there is no strong evidence to choose either a buried charge status or an abnormal pK_a for an acid side chain in these buried pepsinogen aspartic acids. Their influence on activation is unclear. The conserved aspartic acids 87 and 303 are in very similar environments in pepsinogen and pepsin so the effect of these buried carboxyl groups on activation or a different pH stability for precursor and active enzyme is questionable.

Pepsinogen Compared to the Active Enzyme

The most surprising result of the determination of the crystal structure of pepsinogen, first described by James and Sielecki,[17] is the location of the N-terminal residues of the activation peptide in the β-structure at positions occupied in the active enzyme structures by atoms from the N-terminal residues of the enzyme structure. Specifically, when the two structures are superimposed in three dimensions, pepsinogen residues 2P through 9P are structurally equivalent to residues 2 through 9 of pepsin. This structural difference between precursor and active enzyme is illustrated in Figure 9. From residue 14 on toward the C-terminus, the structures of the precursor and active enzyme are very similar. But, whereas residues 1–13 in pepsin precede a turn and are a part of the β-structure, in pepsinogen residues 1–13 extend from the top of the molecule, from the junction of 44P and 1. That junction between the activation peptide and residue 1 of pepsinogen is exposed at the top of the active site cleft. As mentioned before, the activation peptide structure is quite helical and fills the cleft while its N-terminal portion takes part in the β-structure below the active site. Figure 10 demonstrates a portion of pepsinogen which is different after activation. The orientation of the pepsinogen peptide backbone from Glu-13 onward in the C-terminal direction is nearly identical to pepsin and the other active enzyme structures. This is best appreciated by inspecting the orientation of each of the carbonyl groups. The pepsinogen peptide between residues 12 and 13 is very different from that of pepsin in Figure 10. Probably as a consequence of the different course of the backbone and consequential different position for the Glu-13 side chain, the side chain of Phe-15 is in an orientation which is different from the active enzyme structure. The implication of this analysis is that all pepsinogen interactions involving residues from 1P through 13 will be nonexistent (activation peptide) or different (residues 1 through 13) after activation. For example, in pepsinogen the side chain of Glu-13 is involved in ion pair interactions (Table VIII) not only with Arg-8P but also with Arg-307. Neither of these interactions could exist after Glu-13 is transferred to the orientation of the homologous residue of the active enzyme structures. In the pepsin structure Arg-307 ion pairs with Asp-11 and is conserved as arginine or lysine in all known homologous sequences.

The structural adjustments which occur upon pepsinogen activation can be appreciated by examining Figure 11. Most α carbons in the N-terminal lobes of pepsinogen and pepsin superimpose well. In Figure 11, three areas of change are identified in the N-terminal lobe. Glu-7 is the center of the chain segment which undergoes large rearrangement upon activation. Thr-77 identifies the tip of the β-flap, which is nearer the active site in pepsin than in pepsinogen. The side chains of Tyr-75, Thr-77, and Ser-79 contribute to the formation of the S1 binding subsite. They cannot assume their pepsin positions until the activation peptide has been removed. Glu-

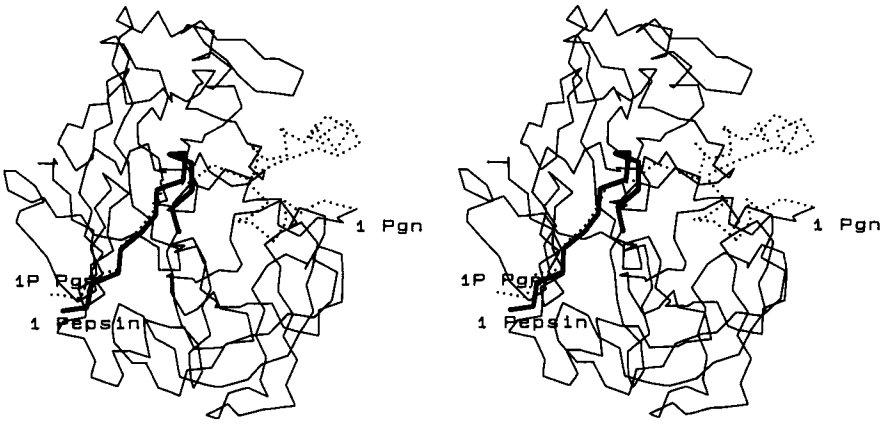

Fig. 9. Pepsinogen α-carbon backbone with backbone from 1 to 15 of pepsin superimposed. The pepsinogen backbone is represented by a dotted line for residues 1P–44P and a thin solid line for residues 1–326. Pepsin residues 1 to 15 are shown as a thick line. The α-carbon positions of Leu-1P and Ile-1 in pepsinogen are labeled as "1P Pgn" and "1 Pgn," respectively. The α-carbon position of Ile-1 in pepsin is labeled as "1 Pepsin." Residues 1P–9P in pepsinogen occupy the same position as residues 1–9 in pepsin.

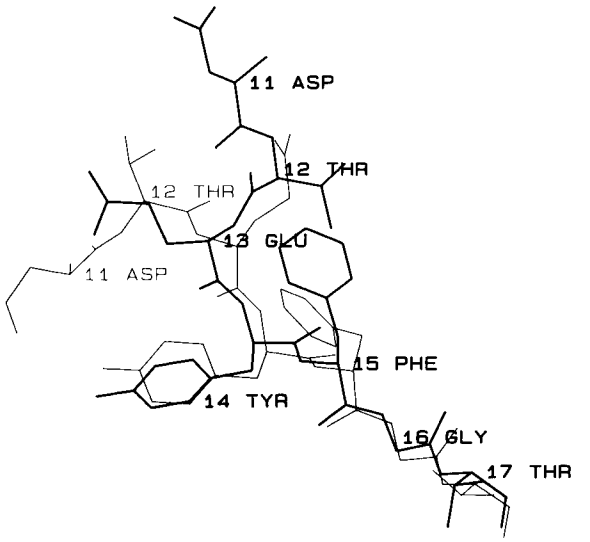

Fig. 10. Comparison of the pepsinogen (heavy line) and pepsin (light line) structures from residues 11 to 17. Pepsinogen residues are labeled (bold) and pepsin residues which do not superimpose (11 and 12) are labeled separately (light). Backbone orientation is very similar in both structures from C_α 13 on toward the C terminus. At residue 13 the backbone of pepsinogen converges with that of pepsin from an origin outside of this view (viz., the 44P–1 junction). By virtue of a rotation about the bond between the C_α and C_β atoms, the side chain of Phe-15 of pepsinogen is different from its conformation in pepsin.

107 identifies a loop which is different in pepsinogen and pepsin. In particular, comparison with pepsin shows that Glu-107 has a significantly different orientation in the two structures. In pepsinogen Glu-107 is exposed on the surface and, in fact, is involved in an intermolecular hydrogen bond in the crystal structure (see below). Although in pepsinogen the side chain of Glu-107 does not interact with activation peptide, residues adjacent in sequence make contact with the N-terminal region of the mature enzyme which relocates upon activation. Consequently, it seems that the Glu-107 rearrangement may be an indirect effect of activation. Gilliland et al.[16] have suggested that Glu-107 (109 in chymosin) is buried upon activation and may be instrumental in the pH induced rearrangements of zymogen activation.

Figure 11 also shows that the alignment of the C-terminal lobes of pepsinogen and pepsin is not nearly so precise as occurs with the N-terminal lobes. The possibility of motion of a subdomain in the C-terminal lobe upon activation has been discussed.[10,11] There are particularly large rearrangements of the regions centered about residues 280 and 292. The wall of the S1′ binding pocket, in pepsin, includes side chains of residues 291, 298, and 300.

Active Site

The active site region of the pepsinogen molecule is shown in Figure 1A. The catalytic groups, Asp-32 and Asp-215, have several interactions with portions of the activation peptide. The ϵ-amino group of Lys-36P is adjacent to both carboxylates in a position which is occupied by water in the active enzyme crystal structures. Also, phenolic oxygens of Tyr-9 and Tyr-37P hydrogen bond to Asp-32 and Asp-215, respectively.

Comparison of the pepsinogen active site to that of pepsin reveals a highly conserved conformation of backbone and side chain atoms with retention of the 2-fold symmetry; indeed, the lack of activity in the zymogen must be attributed to factors other than a distortion of the active site conformation. Alignment of the active site of pepsin to that of pepsinogen,

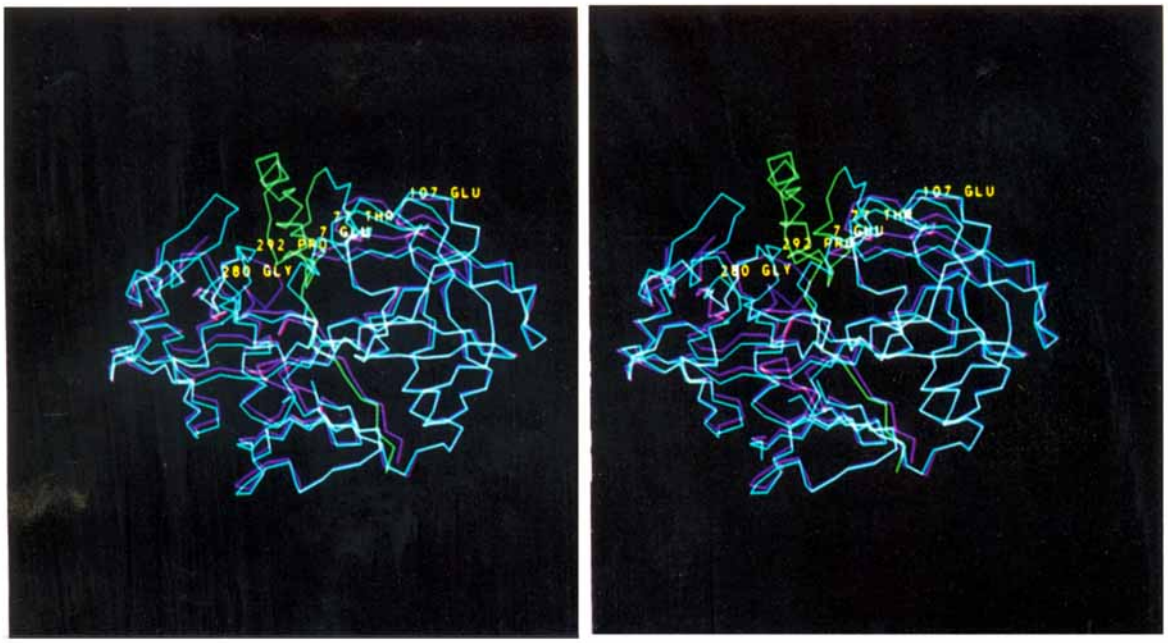

Fig. 11. Stereo view of the superposition of pepsinogen and pepsin α-carbon backbones. Pepsinogen residues 1P–44P are green, and residues 1–326 are blue; pepsin is violet. Gaps in either structure exist when weak electron density prevented coordinate determination. Pepsinogen α-carbon atoms which are mid-points of regions of large structural change are labeled.

using atoms of residues 31–34 and 214–217 (backbone atoms only of residues 31 and 214), gives an rms difference of 0.25 Å. Although solvent is present in the pepsinogen active site and binding cleft, the presence of the activation peptide residues (as well as residues 1–13) precludes solvation to the extent observed in pepsin. Despite the presence of activation peptide and alteration of solvent environment, the active site is conformationally formed in the zymogen.

Conservation of hydrogen bonds involving atoms of the active site likewise indicate the active site conformation is intact to a high degree. Figure 12 depicts the hydrogen bonds in the pepsinogen active site. Hydrogen bonds between the inner† carboxyl oxygen of Asp-32 and backbone nitrogen of Gly-34 are mirrored by Asp-215 and Gly-217 and are preserved in other aspartic proteinase structures. The outer† carboxyl oxygen of Asp-32 accepts a proton from Ser-35 side chain oxygen. However, the symmetrical interaction between Asp-215 and Thr-218 is not observed in pepsinogen, due to a change in the χ_1 dihedral angle of Thr-218, which accepts a hydrogen bond from backbone Leu-220 nitrogen. This symmetrical pair of interactions is present in chymosin, rhizopuspepsin, and endothiapepsin, but in penicillopepsin, and two pepsin structures, the Ser-35 (porcine pepsin numbering) side chain fails to hydrogen bond to Asp-32. Table IX lists the interatomic distances for atoms of the active site aspartic acids in the known gastric and fungal aspartic proteinase structures. The distance alone is not sufficient for defining a hydrogen bond, but Table IX gives an idea of the geometric conservation in an evolutionary cross section of the active site. The fireman's grip[15] (Fig. 12B) is intact in pepsinogen as in other aspartic proteinases. For this characteristic structure, the Thr-33 side chain donates a proton to backbone oxygen of Val-214 and accepts a proton from Thr-216 backbone nitrogen. Symmetrically the side chain of Thr-216 hydrogen bonds with backbone oxygen Phe-31 and backbone nitrogen of Thr-33.

Along with these well-conserved hydrogen bonds, the presence of activation peptide produces additional hydrogen bonding in the active site without distortion of the above mentioned features. The hydroxyl oxygen of Tyr-9 donates its proton to the outer carboxyl oxygen of Asp-32. Note that although Tyr-9 is not part of the activation peptide, its position in the zymogen is not that which it assumes in pepsin (see "Pepsinogen Compared to the Active Enzyme" above). The residues 1–13 perform functions in the zymogen but move to participate in a β-sheet in the mature form. The hydroxyl oxygen of Tyr-37P donates a proton to the outer carboxyl oxygen of Asp-215 as did Tyr-9 to Asp-32. These two residues which show no symmetry in the amino acid sequence come together to form a pair of interactions which obey the 2-fold axis of the active site groups.

†Outer and inner in this section refer to the position of the carboxyl oxygens of Asp-32 and -215 in Figure 12.

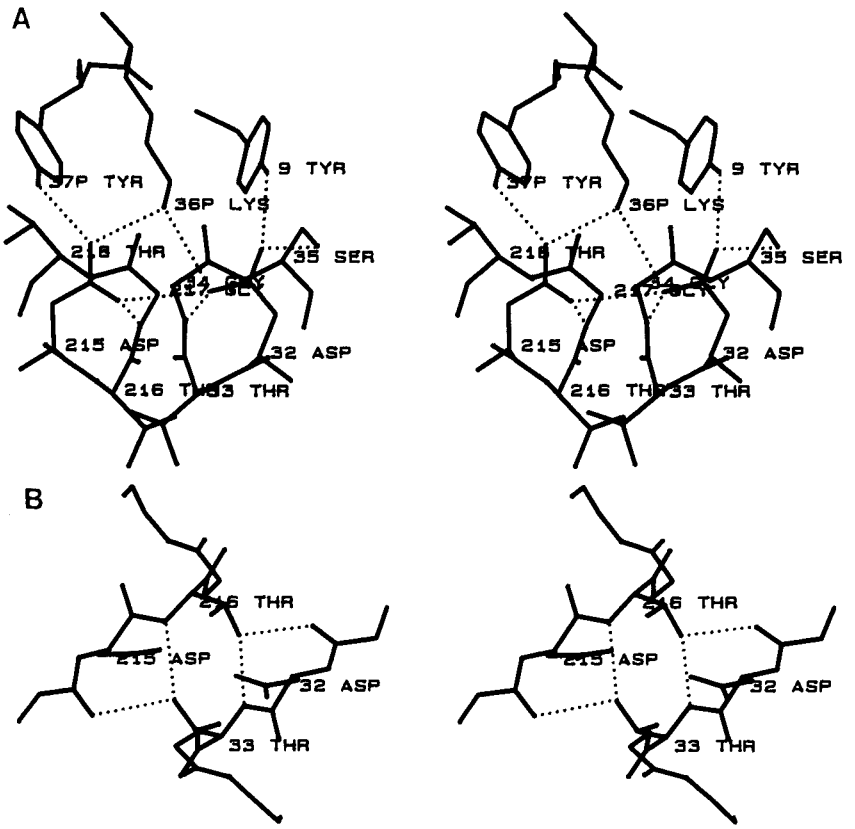

Fig. 12. Hydrogen bonds of the active site. Dotted lines represent hydrogen bonds defined by criteria established by Baker and Hubbard.[42] (**A**) Hydrogen bonds to Asp-32 and -215. (**B**) The conserved Fireman's Grip[15] involving residues 32, 33, 215, and 216.

TABLE IX. Interatomic Parameters for the Active Site of Known Aspartic Proteinases

PDB*	pH	1†	2	3	4	5	6	7	8	9	10	11	12	13
1PSG	6.3	3.5	3.2	2.8	3.5	2.6	2.8	3.4	3.5	6.4	2.5	3.1	−176	66
2PEP	3.6	3.4	2.9	2.9	2.7	3.0	3.2	3.1	3.6	3.1	3.8	3.1	−67	169
3PEP	2.0	3.3	2.9	3.1	3.8	3.1	3.1	3.3	3.4	4.4	2.7	3.2	−67	60
4PEP	2.0	2.9	2.6	3.0	3.0	3.0	3.1	3.5	3.7	3.0	2.8	2.7	−74	73
1CMS	6.0	3.3	3.1	2.9	3.8	2.6	3.1	3.3	3.5	2.9	2.4	3.5	−65	62
4APE	4.5	3.1	2.7	3.0	3.0	2.9	3.3	3.5	3.5	2.8	2.8	2.9	−66	62
2APP	4.4	2.8	2.8	2.9	3.2	2.9	2.8	3.2	3.6	2.8	3.4	2.9	−67	82
2APR	6.0	2.8	2.8	2.9	3.4	2.9	2.9	3.2	3.6	2.6	2.8	3.1	−68	71

*Protein Data Bank[9] entry name: 1PSG, pepsinogen; 2PEP–4PEP, pepsin; 1CMS, chymosin; 4APE, endothiapepsin; 2APP, penicillopepsin; 2APR, rhizopuspepsin.

†Interatomic parameters (distances in Å; residues numbered according to pepsin): 1 Asp-215 inner carboxylate–water (Lys-36P NZ, pepsinogen); 2 Asp-32 inner carboxylate–water (Lys-36P NZ, pepsinogen); 3 Asp-215 outer carboxylate–water (Lys-36P NZ, pepsinogen); 4 Asp-32 outer carboxylate–water (Lys-36P NZ, pepsinogen); 5 Asp-215 inner carboxylate–backbone nitrogen 217; 6 Asp-32 inner carboxylate–backbone nitrogen 34; 7 Asp-215 inner carboxylate–backbone nitrogen 34; 8 Asp-32 inner carboxylate–backbone nitrogen 217; 9 Asp-215 outer carboxylate–side chain oxygen Thr-218; 10 Asp-32 outer carboxylate–side chain oxygen Ser-32; 11 Asp-215 inner carboxylate-Asp-32 inner carboxylate; 12 Thr-218 χ_1 dihedral angle (degrees); 13 Ser-35 χ_1 dihedral angle (degrees).

A superposition of the active site atoms from the C-terminal lobe (backbone of Val-214, backbone and side chain of 215–217) upon their symmetrically related N-terminal atoms is accomplished by a rotation of −179° about the diad axis and gives an rms difference of .71 Å. When this rotation is applied to the hydroxyl oxygen of Tyr-37P the rotated position is only 1.23 Å from the hydroxyl oxygen of Tyr-9. Finally, the ϵ-amino group of Lys-36P is placed medially between and "above" the carboxylate groups of Asp-32 and -215, donating protons to the outer carboxyl oxygen of Asp-215 and the inner of Asp-32.

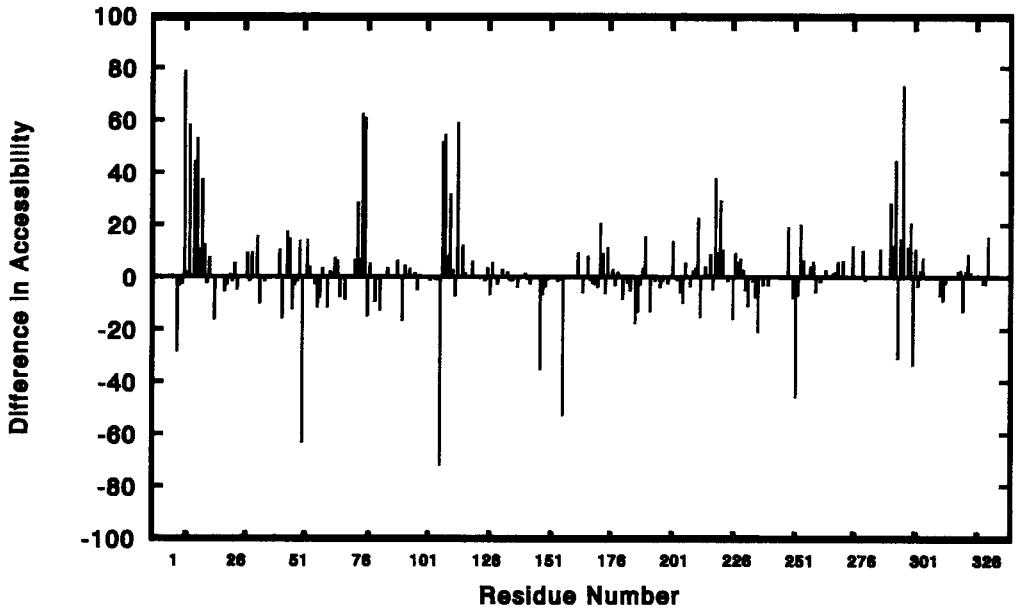

Fig. 13. Difference solvent accessibility plot. The solvent accessibility for a given amino acid, relative to its empirical solvent accessibility in a Gly-X-Gly tripeptide, was computed for pepsinogen and pepsin.[47] The difference accessibility for a given residue is the percentage accessibility in pepsin less that of pepsinogen.

It is possible that the saturation of lone pairs of the outer carboxyl oxygen of Asp-215 with other donor protons makes the acceptance of a hydrogen bond from Thr-218 less favorable. The distance from the outer carboxyl oxygen of Asp-32 to Lys-36P zeta nitrogen is greater than the hydrogen bonding criteria (Table IX), indicating that Asp-32 is not saturated with donor protons to the same extent of Asp-215, thus allowing acceptance of a proton from Ser-35. It cannot be ascertained which of these interactions are causal. A single solvent molecule, Wat-677, is in a position to hydrogen bond to the nitrogen of Lys-36P.

The carboxylate groups of the aspartic acid residues 32 and 215 retain planarity in the zymogen, and a pair of their carboxyl oxygen atoms are within a distance to suggest a possible hydrogen bond between the two, as observed in all other aspartic proteinase crystal structures. Modeling of a proton on the inner carboxyl oxygen of either Asp-32 or -215 indicates the hydrogen bond geometry is more favorable if the proton is assigned to Asp-215. The same exercise, and same conclusion, was made for the other available crystal structures (Table IX). It is interesting that the ϵ-amino group of Lys-36P is in a position similar to that of a solvent molecule in other aspartic proteinase structures and is involved in a similar hydrogen bond network with Asp-32 and -215. The positive charge on Lys-36P suggests the possible presence of a buried anionic charge.

Interactions of the Propeptide With the Active Enzyme

The propeptide, along with residues 1–13, is located in the pepsin substrate binding cleft, with the exception of residues 1P–9P which form a β-strand outside of the cleft. While the N-terminal wall of the cleft is involved with primarily hydrophobic interactions with the propeptide, the C-terminal wall is more or less fully exposed to solvent although there are a few extended side chain to side chain interactions with the propeptide. Figure 13 shows the difference in solvent accessibility (summed over all atoms of a residue) for residues 1–326 of pepsinogen from residues 1–326 of pepsin. Differences of large magnitude do not necessarily reflect the change from buried to accessible or vice versa. Notably, regions 30–35 and 215–219 generally have increased exposure to solvent in pepsin through removal of activation peptide. Residues 1–17 exhibit differences in accessibility due to relocation of the N terminus from a random-coil structure in pepsinogen to the central β-sheet in the pepsin structure. Asp-52, although accessible in both structures, decreases from 88% accessible in pepsinogen to 24% in pepsin, and interacts with Tyr-113. The region 107–120 interacts with the mature N-terminal residues, or with residues of the flap. Residues exhibiting large changes in accessibility which border regions 157–161 or 241–245 are artifacts due to lack of coordinates for those regions in either model. The 287–300

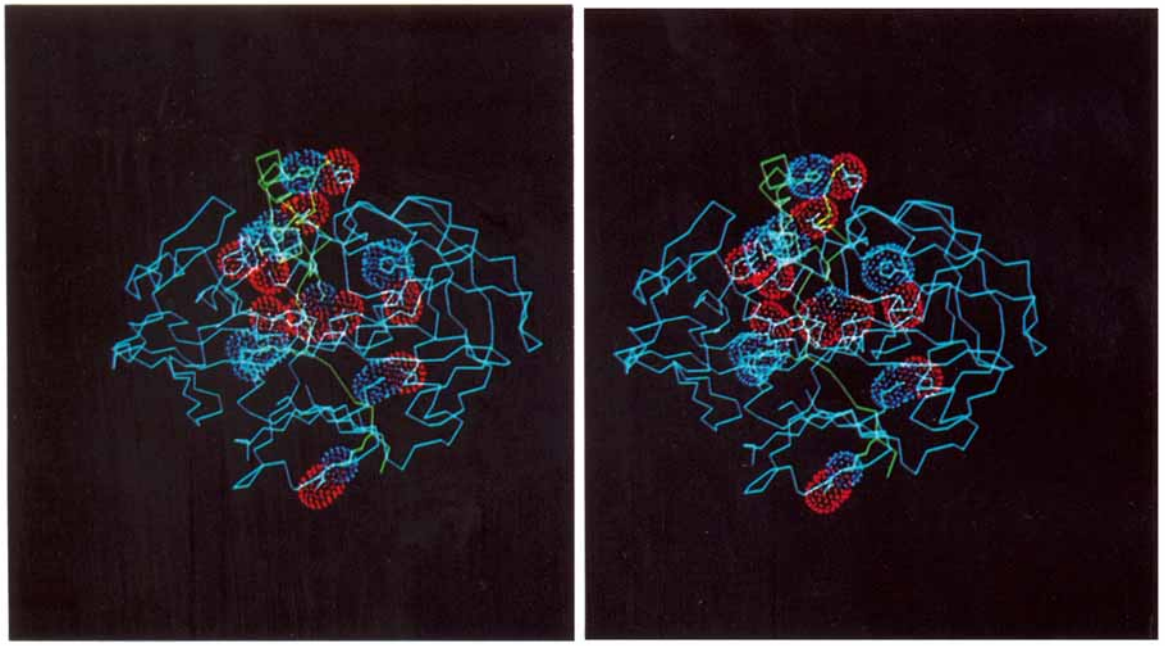

Fig. 14. Distribution of functional groups participating in ion pairs. The green α-carbon backbone is of residues 1P–44P, and the light blue line is of residues 1–326. The side chains of residues involved in ion pairs are the same color as the respective backbone. Anionic or cationic functional groups involved in ion pairs are highlighted by a dot surface, colored red or blue, respectively. The surface is contoured at 2.2 Å. Note the abundance of interactions about the activation peptide.

zone has many changes in accessibility, and interacts directly with activation peptide, as well as with 204–213.

Nine of the 11 ion pairs listed in Table VIII involve the activation peptide or residues expected to have altered positions upon activation. Of those nine, four are between the propeptide and residues 1–13. Figure 14 highlights the charge-bearing functional groups of each ion pair and illustrates that they are localized to the region of the cleft. The aspartate or glutamate side chain of each ion pair is relatively accessible to solvent, with the exception of the catalytic aspartic acids 32 and 215.

The relationship of activation peptide to the substrate binding site was inspected in detail. A consensus of residues which form the pocket of each binding subsite defined in each of the fungal aspartic proteinases[48–52] were tabulated and simultaneously overlaid with the atoms of pepsinogen to determine the position of sidechains involved in subsite formation. Although the active site is conformationally formed in pepsinogen (discussed previously in the "Active Site" section), the binding subsites are distorted. The S1 pocket in Figure 15 is about 3/4 intact in pepsinogen. The remainder of the S1 pocket, formed by residues of the flap and Phe-112 in the active enzyme–inhibitor structures, in pepsinogen is formed by backbone and side chain of residues 6–8. The side chain of Tyr-9 occupies the subsite. The Sl′ pocket is likewise mostly formed, and the side chain of Tyr-37P fills this pocket. Some residues which defined S1′ pocket in the active proteinase–inhibitor structures show a χ_1 or χ_2 shift to accommodate Tyr-37P while two others are completely replaced by Ala-34P and Phe-38P. The other subsites were not as well defined in the inhibitor complexes as were S1 and S1′. Moreover, the S2, S3, S2′, and S3′ subsites show distortion due largely to the presence of activation peptide, and the voids of these pockets are not as obviously occupied with side chains from the activation peptide, as are the pockets of S1 and S1′.

Intermolecular Contacts in the Crystal

Consideration of the other pepsinogen molecules present in the crystal of space group $C2$ reveals several intermolecular interactions, including one ion pair and several hydrogen bonds, some of which are solvent mediated. Intermolecular hydrogen bonds listed in Table X were determined according to parameters defined by Baker and Hubbard.[42] Of those 12 hydrogen bonds only four are side chain to side chain, indicative of areas of close packing in the crystal. The single intermolecular ion pair pictured in Figure 16 occurs between the N-terminus and OD1 of Asp-171 from a molecule related by the crystal diad at $x = 1/2$ and $z = 0$. Also, the leucine side chain of residue 1P has a van der Waals interaction across the same diad to the leucine side chain of residue 1P in the symmetry-related molecule. In addition to forming an ion pair with the N-terminus of a symmetry related molecule, the Asp-171 side

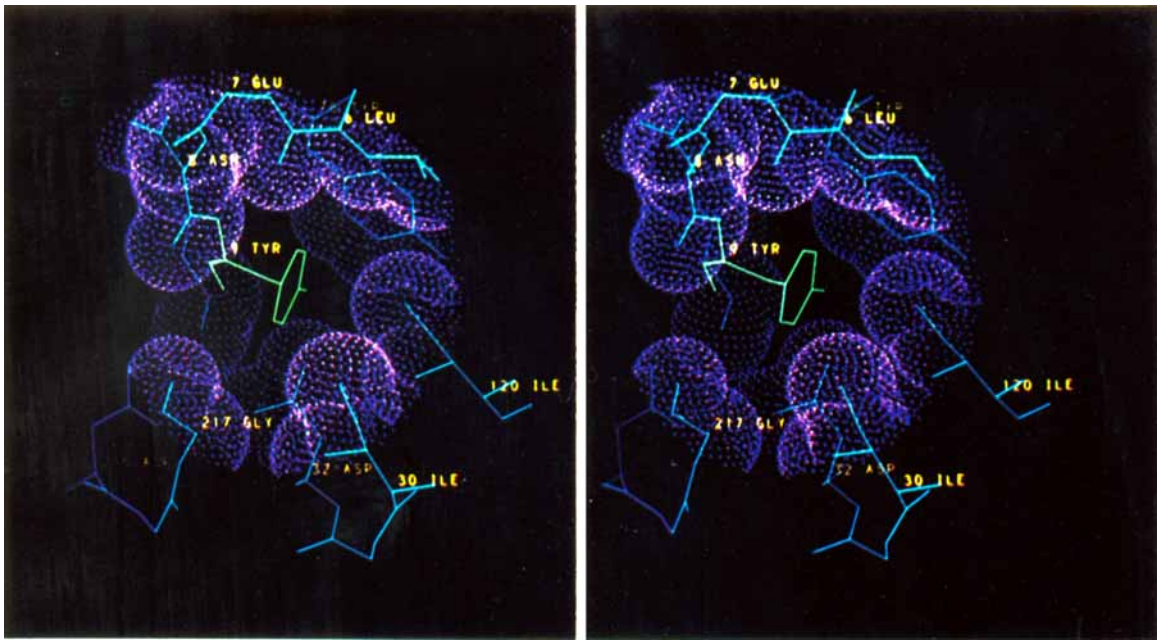

Fig. 15. Residues defining the S1 binding subsite of pepsinogen. Residues pictured in blue which are included in the S1 subsite were determined according to the procedure described in the text. van der Waals dot surface is pictured for atoms of the S1 subsite which contact Tyr-9 (green) within a 5.0 Å cutoff.

chain also makes an intramolecular ion pair with the Lys-3P side chain (Table VIII). A computation of accessibility[47] which included the atoms of neighboring molecules revealed that the residues involved in intermolecular contacts have a reduced solvent accessibility. However, only four of those side chains are completely buried in the crystal, including Phe-25P, Ala-49, Ile-170, and Ala-227. Many hydrophobic residues located on the surface have drastically reduced surface area but are not completely buried as a result of intermolecular contacts.

Another interesting intermolecular interaction involves the Glu-107 side chain which is buried in the pepsin molecule. In pepsinogen the side chain of Glu-4 is within a distance to hydrogen bond to the side chain of Glu-107 of symmetry-related molecule (Table X). The side chain of Glu-107 is 74% accessible in a single molecule of pepsinogen, and 13% accessible when surface area is calculated in the presence of neighboring molecules. The side chain of Glu-4 is likewise 7 and 4% accessible, respectively. The reduced accessibility of this pair by virtue of the crystal contact may be sufficient to shield one of the carboxyl groups from ionization, allowing for a carboxylate–carboxylate hydrogen bond. (A similar bond is observed in the active site between aspartic acids 32 and 215.)

DISCUSSION

Knowledge of the pepsinogen crystal structure allows a new depth of understanding of the activation process of the aspartic proteinases. The novel aspect of this activation is that it may occur unimolecularly with the only stimulus being a lowering of the pH of the environment of the molecule. The activation peptide, generally speaking, occupies the nascent binding cleft of the enzyme. Its secondary structure is primarily helical and interactions with the pepsin portion of the molecule are limited. However, one stretch of chain, residues 2P through 6P, participates in β-structure by forming hydrogen bonds to residues 163 through 167. Comparison with the pepsin structure demonstrates that this activation peptide chain is replaced by the N-terminal chain of the mature enzyme. Consequently, during activation this β-structure is disrupted, and after conformational change, new β-structure is formed. The detailed structure of pepsinogen restricts the spectrum of possible mechanisms of activation on the basis of steric considerations. Examination of the structure, in detail, suggests certain interactions which may be critical to the stabilization of the pepsinogen structure and which must be destroyed for activation to occur. Results presented here are in agreement with those from James and Sielecki.[17]

As is described in the "Active Site" section above, the significant hydrogen bonding in the zymogen nascent active site is the Tyr-37P, Lys-36P, Tyr-9 triad positioned in the nascent substrate binding area to interact with the active site aspartyls. These residues are not a pseudo-substrate since the side chains not a peptide are directly interacting with the aspartic acid residues. Surprisingly, compari-

TABLE X. Intermolecular Hydrogen Bonds

Atom of molecule *x,y,z*	Atom of symmetry related molecule	Crystal symmetry transformation			Distance (Å)
Leu-1P N	Asp-171 OD1	$1-x$	y	$-z$	2.70
Asp-3 OD1	Glu-107 N	$1-x$	y	$1-z$	3.13
Glu-4 OE2	Glu-107 OE2	$1-x$	y	1−z	2.94
Leu-48 N	Ser-110 OG	$1-x$	y	$1-z$	3.04
Asp-59 O	Ser-226 OG	$1/2+x$	$1/2+y$	z	3.29
Glu-65 OE1	Cys-249 N	$1/2+x$	$1/2+y$	z	2.69
Glu-65 OE2	Ser-250 N	$1/2+x$	$1/2+y$	z	2.83
Glu-65 OE2	Ser-250 OG	$1/2+x$	$1/2+$y	z	2.63
Glu-107 OE2	Ser-110 N	$1-x$	y	$1-z$	3.04
Glu-107 OE2	Ser-178 OG	$1-x$	y	1−z	2.51
Gly-144 O	Ser-178 OG	$1-x$	y	−z	3.04
Ser-147 OG	Tyr-175 OH	1−x	y	−z	2.89

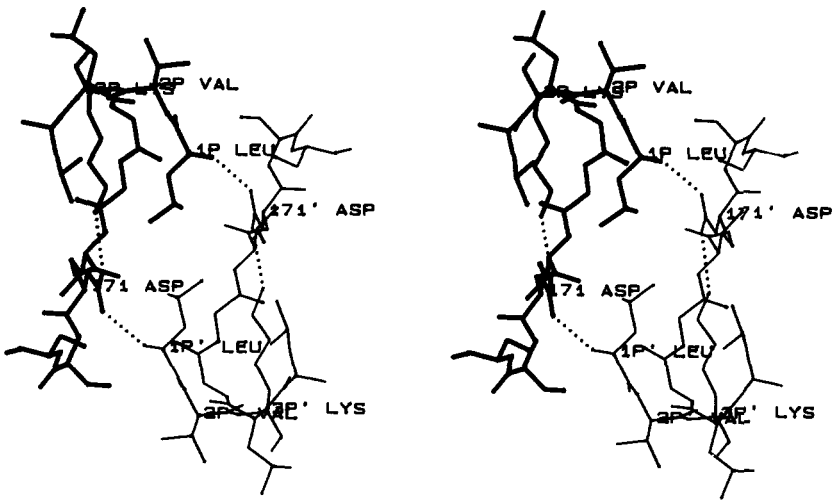

Fig. 16. Intermolecular contacts in the region of the N-terminus. Molecule *x,y,z* is pictured with bold lines, and molecule $1-x,y,-z$ is pictured with thin lines. Dotted lines are hydrogen bonds or ion pairs. Note that the ion pairs of Asp-171 and Lys-3P and the N-terminus of the neighboring molecule both qualify as hydrogen bonds.

sons with crystal structures of mature aspartic proteinases reveal that even though the pepsinogen active site has the extra charged amino group of Lys-36P, there is no evidence in the pepsinogen crystal structure for an altered orientation of the aspartic acids or for a change in the distance between the carboxyl groups. Consequently, there is no structural evidence for a change in the state of protonation of the active site residues upon activation. Of course, the lack of activity of pepsinogen is explained by the inaccessibility of the active site.

Since the structure of the pepsin portion of the pepsinogen molecule is virtually identical to the structures of the known active aspartyl proteinases, it seems logical to assume that the pepsin portion of the pepsinogen molecule does not undergo large conformational changes during the dynamic activation process. Consequently, the interactions most likely to be disturbed during activation are those within the activation peptide and between the activation peptide and the pepsin portion of the molecule. Because the stability of the ion pairs will be altered upon a drop of pH and consequent discharge of the carboxylic acids, attention is directed first toward these interactions. As shown in Table VIII, 9 of the 11 ion pairs in pepsinogen will not be present in pepsin because, in each case, the positive moiety is contained in the activation peptide. With the exception of the two interactions made by Lys-36P and the active site aspartyl groups, the other nine ion pairs are accessible to solvent and would be expected to respond readily to a change in pH. There are five ion pairs in which the positive moiety is part of the activation peptide and the carboxylic acid moiety is in the part of pepsin which must undergo conformational change in order to assume its pepsin configuration viz. 3, 7, 11, or 13 (see Fig. 9). Finally, there are three ion pairs between positive charges in the

activation peptide and acidic side chains which have essentially the same configuration in both pepsin and pepsinogen.

The hydrophobic side chains of the activation peptide have interactions with one another as well as in some instances with hydrophobic areas from the pepsin portion of the molecule. Specifically, the N-terminal peptide (1P through 6P) which is in the pepsin β-structure interacts with pepsin residues 165, 166, and 167 of the adjacent β-chain as well as with residues 14, 31, and 94. Residues 37P and 38P have hydrophobic interactions with pepsin residues 189, 213, 291, 298, and 300. This is the substantially formed S1′ binding pocket of the enzyme. There is a small hydrophobic core within the activation peptide which involves residues 12P, 16P, 17P, 22P, 25P, and 26P. Only Phe-111 and Tyr-114 contact this activation peptide hydrophobic core. In pepsinogen, residues 6 and 9 have hydrophobic interactions with pepsin residues 30, 39, 75, 117, and 120. Tyr-9 occupies the S1 binding pocket. After activation, the backbone which includes residues 6 and 9 has undergone rearrangement so that different contacts are made. The consequences of removal of the activation peptide on the hydrophobic structure of pepsin would be minimal. The interactions involved in the β-structure will be replaced after activation by similar interactions with the N-terminus of pepsin. The binding pockets will be unoccupied.

The pepsinogen structure allows several conclusions concerning the activation of the zymogen. First, most straightforward is the understanding of the bimolecular activation by preformed pepsin. Pepsin results from cleavage between Leu-44P and Ile-1. This bond is one that the substrate specificity of pepsin predicts would be readily cleaved by pepsin. Moreover, the loop containing the bond to be cleaved is superficial. In fact, in the pepsinogen electron density maps, there is not sufficient density to allow assignment of positions for the side chain atoms of residues 42P through 1. This weak density is the result of unusually high flexibility of the protein in this portion of the molecule. Most likely when pepsinogen is activated by another preformed pepsin molecule, the cleavage at 44P-1 occurs initially; then, the activation peptide is degraded further.

The course of unimolecular pepsinogen activation is more difficult to predict. There are chemical experiments which show that the first (unimolecular) cleavage may occur between residues 16P and 17P[53,54] or between residues 44P and 1.[8] Since the most dramatic effect of activation is the removal of residues 2P through 6P from the β-structure and replacement by residues 2 through 6, it is appropriate to ask if the structure dictates any sequential relationship to the three events viz. removal of the chain from the β-structure, cleavage of the peptide bond, and insertion of the 1 through 6 chain in the β-structure. In the pepsinogen structure, the active site is occupied by the side chains of residues Lys-36P, Tyr-37P, and Tyr-9 so that no peptide could approach the active site carboxyl groups. Motion of these residues must occur before unimolecular activation could take place. However, from pure distance considerations, the β-structure interactions of 2P through 9P could remain intact, while either 16P–17P or 44P-1 is positioned in the active site. Consequently, the structure of pepsinogen does not provide enough information to define the order of the steps of unimolecular activation. The suggestion of alternative positions of the amino terminal residues (1P–9P) of the activation peptide during unimolecular activation has been made previously.[17] It does seem unlikely that residues 2 through 6 are placed in their pepsin positions before cleavage of the 44P-1 bond because the N-terminus of pepsin is far removed from the binding cleft. If this were to occur, the motion required of the activation peptide would be so great that it is unlikely.

The pepsinogen structure does not dictate a choice for unimolecular bond cleavage at either 16P–17P or 44P-1. Both of these peptide bonds are near the outer surface of the binding cleft and would have to be repositioned before either bond could by lysed. Residues 16P and 17P are part of a helix which would have to be disrupted before the scissile bond could fit in the active site of the nascent pepsin.

The Pro-Leu dipeptide of pepsinogen residues 5P and 6P participates in β-structure and is replaced by pepsin residues Pro-5 and Leu-6 upon activation. This sequence identity suggests that these side chains may target the rearranging pepsin N-terminal chain to its proper pepsin position. Figure 17 shows a sequence alignment for the known aspartic protease propeptides. For the other enzymes shown in Figure 17, the structural replacements are as follows: pepsin C, PL by PL; chymosin, PL by PL; renin, FL by VL; cathepsin D, PL by VL; proteinase A, KI by PL; and rhizopuspepsin, PG by PM. This high degree of sequence homology lends credence to the targeting suggestion.

Since porcine pepsinogen is the only aspartic protease zymogen whose crystal structure is known, it is appropriate to consider whether pepsinogen is a good model for the other zymogens. Figure 17 shows a sequence alignment for the known zymogens. In making this alignment, the tertiary structure of pepsinogen has been used as a guide to determine good places for deletions. Specifically, the four residue deletion in proteinase A was placed so that one turn at the end of a helix was deleted. Examination of this figure demonstrates that the side chains which interact with the active site, viz. Lys-36P and Tyr-37P, are conserved in the other enzymes. The only exception is renin where the closest match is Lys-Arg rather than Lys-Tyr. In only one instance is a buried pepsinogen residue replaced by a charged side chain. At position 22P, leucine in pepsinogen is

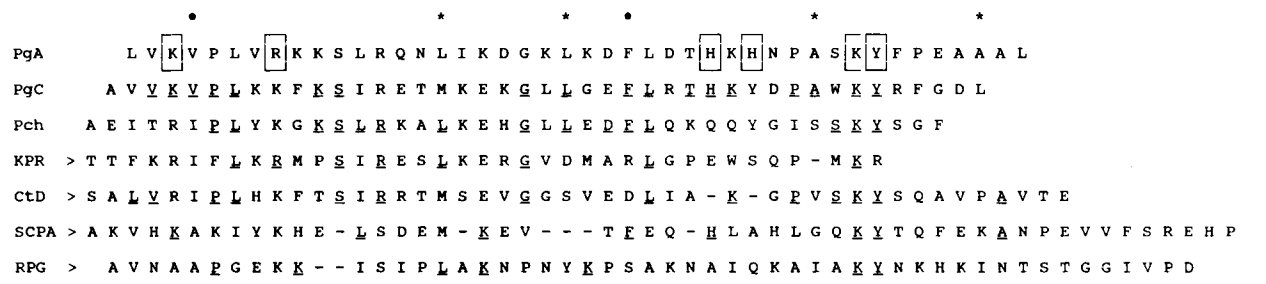

Fig. 17. Alignment of amino acid sequence of activation peptides from various aspartic proteinases.[3,6] Knowledge of structural elements of porcine pepsinogen was used to guide the alignment. Abbreviations: PgA, porcine pepsinogen A; PgC, human pepsinogen C; Pch, bovine prochymosin; KPR, human kidney prorenin; CtD, human procathepsin D; SCPA, proteinase A; RPG, rhizopuspepsinogen. A > preceding a sequence indicates the sequence is inferred from nucleotide sequencing. In PgA boxed amino acids are structurally important, i.e., participate in ion pairs or hydrogen bond to the active site carboxylates. Asterisks above the PgA sequence signify residues which are inaccessible to solvent. Underlined amino acids are identical to those in PgA.

replaced by aspartic acid in renin. In rhizopuspepsinogen, proline occurs at positions 15P and 24P, both of which are within helices in pepsinogen. Acknowledging that adjustments in the structures would have to be made to accommodate the above mentioned variations, in general a high degree of structural homology between pepsinogen and the other aspartic proteinase zymogens seems likely.

Note: Following submission of this manuscript, a detailed description by Sielecki et al.[56] of an independent crystal structure determination of porcine pepsinogen appeared in the literature and shows no significant differences from the present work.

ACKNOWLEDGMENTS

We are indebted to Drs. S.N. Rao and Stanley Koszelak for the initial data collection for pepsinogen and the heavy atom derivative.

REFERENCES

1. Kay, J. Aspartic proteinases and their inhibitors. In: Aspartic Proteinases and Their Inhibitors. Kostka, V., ed. Berlin: Walter de Gruyter, 1985: 1–18.
2. Davies, D.R. The structure and function of aspartic proteinases. Annu. Rev. Biophys. Biophys. Chem. 19:189–215, 1990.
3. Tang, J., Wong, R.N.S. Evolution in the structure and function of aspartic proteases. J. Cell. Biochem. 33:53–63, 1987.
4. Foltmann, B. Gastric proteinases-structure, function, evolution and mechanism of action. Essays Biochem. 17:52–84, 1981.
5. Higashimori, K., Marcott, P.A., Eagan, D.A., Holleman, W.H., Heusser, C., Poisner, A.M., Inagami, T. Pure human inactive renin. J. Biol. Chem. 264:14662–14667, 1989.
6. Chen, Z., Koelsch, G., Han, H., Wang, X., Lin, X., Hartsuck, J.A., Tang, J. Recombinant rhizopuspepsinogen. Expression, purification, and activation properties of recombinant rhizopuspepsinogens. J. Biol. Chem. 266: 11718–11725, 1991.
7. Al-Janabi, J., Hartsuck, J.A., Tang, J. Kinetics and mechanism of pepsinogen activation. J. Biol. Chem. 247:4628–4632, 1972.
8. Kageyama, T., Takahashi, K. Activation mechanism of monkey and porcine pepsinogens A. Eur. J. Biochem. 165: 483–490, 1987.
9. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rogers, 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.
10. Cooper, J.B., Khan, G., Taylor, G., Tickle, I.J., Blundell, T.L. X-ray analyses of aspartic proteinases II. Three-dimensional structure of the hexagonal crystal form of porcine pepsin at 2.3 Å resolution. J. Mol. Biol. 214:199–222, 1990.
11. Abad-Zapatero, C., Rydel, T.J., Erickson, J. Revised 2.3 Å structure of porcine pepsin: Evidence for a flexible subdomain. Proteins 8:62–81, 1990.
12. Sielecki, A.R., Fedorov, A.A., Boodhoo, A., Andreeva, N.S., James, M.N.G. Molecular and crystal structures of monoclinic porcine pepsin refined at 1.8 Å resolution. J. Mol. Biol. 214:143–170, 1990.
13. 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.
14. James, M.N.G., Sielecki, A.R. Structure and refinement of penicillopepsin at 1.8 Å resolution. J. Mol. Biol. 163:299–361, 1983.
15. Pearl, L., Blundell, T. The active site of aspartic proteinases. FEBS Lett. 174:96–101, 1984.
16. Gilliland, G.L., Winborne, E.L., Nachman, J., Wlodawer, A. The three-dimensional structure of recombinant bovine chymosin at 2.3 Å resolution. Proteins 8:82–101, 1990.
17. James, M.N.G., Sielecki, A.R. Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8 Å resolution. Nature (London) 319:33–38, 1986.
18. Rao, S.N., Koszelak, S.N., Hartsuck, J.A. Crystallization and preliminary crystal data of porcine pepsinogen. J. Biol. Chem. 252:8728–8730, 1977.
19. Reeke, G.N. The ROCKS system of computer programs for macromolecular crystallography. J. Appl. Crystallogr. 17: 125–130, 1984.
20. Matthews, B.W., Czerwinski, E.W. Local Scaling: A method to reduce systematic errors in isomorphous replacement and anomalous scattering measurements. Acta Crystallogr. A31:480–487, 1975.
21. Schmid, M.F., Weaver, L.H., Holmes, M.A., Grutter, M.G., Ohlendorf, D.H., Reynolds, R.A., Remington, S.J., Matthews, B.W. An oscillation data collection system for high-resolution protein crystallography. Acta Crystallogr. A37: 701–710, 1981.
22. Crowther, R.A. The fast rotation function. In: The Molecular Replacement Method. Rossman, M.G., ed. New York: Gordon and Breach, 1972: 173–178.
23. Vyas, N.K., Vyas, M.N., Quiocho, F.A. The 3 Å resolution structure of a D-galactose-binding protein for transport and chemotaxis in *E. coli*. Proc. Natl. Acad. Sci. U.S.A. 80:1792–1796, 1983.
24. Crowther, R.A., Blow, D.M. A method of positioning a known molecule in an unknown crystal structure. Acta Crystallogr. 23:544–548, 1967.
25. Hendrickson, W.A., Lattman, E.E. Representation of phase probability distributions for simplified combination of independent phase information. Acta Crystallogr. B26: 136–143, 1970.
26. Sim, G.A. The distribution of phase angles for structures containing heavy atoms. II. A modification of the normal

heavy-atom method for non centrosymmetric structures. Acta Crystallogr. 12:813–815, 1959.
27. Reynolds, R.A., Remington, S.J., Weaver, L.H., Fisher, R.G., Anderson, W.F., Ammon, H.L., Matthews, B.W. Structure of a serine protease from rat mast cells determined from twinned crystals by isomorphous and molecular replacement. Acta Crystallogr. B41:139–147, 1985.
28. Remington, S.J., Woodbury, R.J., Reynolds, R.A., Matthews, B.W., Neurath, H. The structure of rat mast cell protease II at 1.9 Å resolution. Biochemistry 21: 8097–8105, 1988.
29. Cohen, G.H., Silverton, E.W., Davies, D.R. Refined crystal structure of gamma-chymotrypsin at 1.9 Å: Comparison with other pancreatic serine proteases. J. Mol. Biol. 148: 449–479, 1981.
30. Tsukada, H., Blow, D.M. Structure of alpha-chymotrypsin refined at 1.68 Angstrom resolution. J. Mol. Biol. 184:703–711, 1985.
31. Meyer, E.F., Jr., Radhakrishnan, G.M.C., Presta, L.G. Structure of the product complex of acetyl-Ala-Pro-Ala with porcine pancreatic elastase at 1.65 Angstrom resolution. J. Mol. Biol. 189:533–539, 1986.
32. Tronrud, D.E., Ten Eyck, L.F., Matthews, B.W. An efficient general-purpose least-squares refinement program or macromolecular structures. Acta Crystallogr. A43:489–501, 1987.
33. Jones, T.A. A graphics model building and refinement system for macromolecules. J. Appl. Crystallogr. 11:268–272, 1978.
34. Jack, A., Levitt, M. Refinement of large structures by simultaneous minimization of energy and R factor. Acta Crystallogr. A34:931–935, 1978.
35. Remington, S.J., Wiegand, G., Huber, R. Cyrstallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 Angstrom resolution. J. Mol. Biol. 158:111–152, 1982.
36. 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.
37. Luzzati, V. Traitement statiztique des erreurs dans la determination des structures cristallines. Acta Crystallogr. 5:802–810, 1952.
38. Ramachandran, C., Ramachandran, G.N. Stereochemical criteria for polypeptide and protein chain conformations II. Allowed conformations for a pair of peptide units. Biophys. J. 5:909–933, 1965.
39. 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.
40. Andreeva, N.S., Gustchina, A.E. On the supersecondary structure of acid proteases. Biochem. Biophys. Res. Commun. 87:32–42, 1979.
41. Blundell, T.L., Sewell, B.T., McLachlan, A.D. Four-fold structural repeat in the acid proteases. Biochim. Biophys. Acta 580:24–31, 1979.
42. Baker, E.N., Hubbard, R.E. Hydrogen bonding in globular proteins. Prog. Biophys. Mol. Biol. 44:97–179, 1984.
43. Barlow, D.J., Thornton, J.M. Helix geometry in proteins. J. Mol. Biol. 201:601–619, 1988.
44. Richardson, J.S. Protein anatomy. Adv. Protein Chem. 34: 167–339, 1981.
45. Barlow, D.J., Thornton, J.M. Ion-pairs in proteins. J. Mol. Biol. 168:867–885, 1983.
46. Foltmann, B. Aspartic proteinases: alignment of sequences. In Abstr. 18th Linderstrom-Lang Conf. 1988: 7–20.
47. Lee, B., Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. 55: 379–400, 1971.
48. James, M.N.G., Sielecki, A.R., 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.
49. James, M.N.G. Sielecki, A.R. Stereochemical analysis of peptide bond hydrolysis catalyzed by the aspartic proteinase penicillopepsin. Biochemistry 24:3701–3713, 1985.
50. Cooper, J., Foundling, S., Hemmings, A., Blundell, T., Jones, D.M., Hallett, A., Szelke, M. The structure of a synthetic pepsin inhibitor complexed with endothiapepsin. Eur. J. Biochem. 169:215–221, 1987.
51. Foundling, S.I., Cooper, J., Watson, F.E., Pearl, L.H., Hemmings, A., Wood, S.P., Blundell, T., Hallett, A., Jones, D.M., Sueiras, J., Atrash, B., Szelke, M. Cyrstallographic studies of reduced bond inhibitors complexed with an aspartic proteinase. J. Cardiovasc. Pharamcol. 10:S59–S68, 1987.
52. Hofmann, T., Hodges, R.S., James, M.N.G. Effect of pH on the activities of penicillopepsin and rhizopuspepsin and a proposal for the productive substrate binding mode in penicillopepsin. Biochemistry 23:635–643, 1984.
53. Dykes, C.W., Kay, J. Conversion of pepsinogen into pepsin is not a one-step process. Biochem. J. 153:141–144, 1976.
54. Christensen, K.A., Pedersen, V.B., Foltmann, B. Identification of an enzymatically active intermediate in the activation of porcine pepsinogen. FEBS Lett. 76:214–218, 1977.
55. Hsu, I.N., Delbaere, L.T.J., James, M.N.G. Penicillopepsin from *Penicillium janthinellum* crystal structure at 2.8 Å and sequence homology with porcine pepsin. Nature (London) 266:140–145, 1977.
56. Sielecki, A.R., Fujinaga, M., Read, R.J., James, M.N.G. Refined structure of porcine pepsinogen at 1.8 Å resolution. J. Mol. Biol. 219:671–692, 1991.