Refined Atomic Model of Wheat Serine Carboxypeptidase II at 2.2-Å Resolution†,‡

Der-Ing Liao,§,∥ Klaus Breddam,⊥ Robert M. Sweet,# Timothy Bullock,∥ and S. James Remington*,∥

Institute of Molecular Biology and Department of Physics, University of Oregon, Eugene, Oregon 97403, Department of Chemistry, Carlsberg Laboratory, Gamle Carlsbergvej 10, DK-2500 Copenhagen Valby, and Biology Department, Brookhaven National Laboratory, Upton, Long Island, New York 11973

Received June 9, 1992; Revised Manuscript Received July 22, 1992

ABSTRACT: The crystal structure of the homodimeric serine carboxypeptidase II from wheat (CPDW-II, M_r 120K) has been determined and fully refined at 2.2-Å resolution to a standard crystallographic R factor of 16.9% using synchrotron data collected at the Brookhaven National Laboratory. The model has an rms deviation from ideal bond lengths of 0.018 Å and from bond angles of 2.8°. The model supports the general conclusions of an earlier study at 3.5-Å resolution and will form the basis for investigation into substrate binding and mechanistic studies. The enzyme has an $\alpha+\beta$ fold, consisting of a central 11-stranded β-sheet with a total of 15 helices on either side. The enzyme, like other serine proteinases, contains a "catalytic triad" Ser146–His397–Asp338 and a presumed "oxyanion hole" consisting of the backbone amides of Tyr147 and Gly53. The carboxylate of Asp338 and imidazole of His397 are not coplanar in contrast to the other serine proteinases. A comparison of the active site features of the three families of serine proteinases suggests that the "catalytic triad" should actually be regarded as two diads, a His–Asp diad and a His–Ser diad, and that the relative orientation of one diad with respect to the other is not particularly important. Four active site residues (52, 53, 65, and 146) have unfavorable backbone conformations but have well-defined electron density, suggesting that there is some strain in the active site region. The binding of the free amino acid arginine has been analyzed by difference Fourier methods, locating the binding site for the C-terminal carboxylate of the leaving group. The carboxylate makes hydrogen bonds to Glu145, Asn51, and the amide of Gly52. The carboxylate of Glu145 also makes a hydrogen bond with that of Glu65, suggesting that one or both may be protonated. Thus, the loss of peptidase activity at pH > 7 may in part be due to deprotonation of Glu145. The active site does not reveal exposed peptide amides and carbonyl oxygen atoms that could interact with substrate in an extended β-sheet fashion. The fold of the polypeptide backbone is completely different than that of trypsin or subtilisin, suggesting that this is a third example of convergent molecular evolution to a common enzymatic activity. Furthermore, it is suggested that the active site sequence motif "G–X–S–X–G/A", often considered the hallmark of serine peptidase or esterase activity, is fortuitous and not the result of divergent evolution. However, the fold of CPDW-II has a remarkable similarity to four other classes of hydrolytic enzymes for which representative structures have recently been determined. These are the acetylcholine esterase, several triacylglycerol lipases, haloalkane dehalogenase, and dienelactone hydrolase. Thus, serine carboxypeptidases are suggested to illustrate the process of both convergent and divergent molecular evolution. A comparison of the amino acid sequences of human serine carboxypeptidase carbL, or "protective protein", and CPDW-II suggests that the dimer of CPDW-II may be a reasonable model for the dimer of the human enzyme and that the dimer interface is similar in the two proteins. This may be useful toward the understanding of human genetic disorders such as galactosialidosis, which can result from lesions in this locus [Zhou, X. Y., Galjart, N. J., Willemsen, R., Gillemans, N., Galjaard, H., & d'Azzo, A. (1991) *EMBO J. 10,* 4041–4048].

Serine carboxypeptidases have been found in virtually every higher organism examined and have a variety of tissue specific and cell compartment specific functions, some of which have only recently been elucidated. Serine carboxypeptidases from many sources have been characterized and are usually glycoproteins with subunit molecular mass of 50–70 kDa and form a variety of oligomeric states. For an extensive review of the properties of this enzyme from several sources, see Breddam (1986).

Serine carboxypeptidases are serine proteinases and, like trypsin- or subtilisin-like enzymes, have an active site seryl residues that forms a covalent adduct with diisopropylphosphofluoridate (Hayashi et al., 1973), causing irreversible loss of activity. Liao and Remington (1990) have shown that the enzyme from wheat has a catalytic triad comprised of Asp338–His397–Ser146 that is similar in arrangement to those of the trypsin or subtilisin-like enzymes. The enzyme will hydrolyze esters and also deamidate amino acid amides. However, unlike the trypsin- or subtilisin-like enzymes, the peptidase activity has a pH optimum in the range 4.5–5.5, depending on the enzyme, and is reduced by more than 90% above pH 7. It has been shown (Breddam, 1985; Breddam et al., 1987) that low activity toward peptide substrates at high pH results from an increase in K_m, not a decrease in k_{cat}, since activity toward ester substrates is maximal above pH 7. This suggests that

† This work was supported in part by grants from the National Science Foundation (DMB 8817438 and MCB 9118302 to S.J.R.) and a grant to the Institute of Molecular Biology from the Lucille P. Markey Charitable Trust.

‡ The coordinates have been deposited in the Brookhaven Protein Data Bank (3SC2).

* To whom correspondence should be addressed.

§ Present address: Center for Advanced Research in Biotechnology, Maryland Biotechnology Institute, University of Maryland, 9600 Gudelsky Drive, Rockville, MD 20850.

∥ University of Oregon.

⊥ Carlsberg Laboratory.

Brookhaven National Laboratory.

0006-2960/92/0431-9796$03.00/0

the deprotonation of an active-site residue involved in peptide substrate binding results in the lowered affinity toward peptides at high pH.

Interest in the activity of serine carboxypeptidases is rapidly increasing for a variety of reasons. For example [Steiner et al., 1980; reviewed by Douglass et al. (1984)], *in Saccharomyces cerevisiae* it has been demonstrated that the maturation of the α-mating pheromone involves a two-step procedure, endoproteolysis by the product of the *KEX2* gene at the carboxyl side of a pair of basic residues, followed by two-step removal of the basic pair by the membrane-bound product of the *KEX1* gene, which is a serine carboxypeptidase (Thomas et al., 1990; Cooper & Bussey, 1989). Thomas et al. (1989) have shown that expression of these two yeast genes in a BSC-40 cell line along with the cDNA for mouse proopiomelanocortin (mPOMC) caused full conversion of mPOMC to mature neuropeptides, including γ-lipotropin. Therefore, there is a significant functional similarity of the peptide hormone maturation machinery of these diverse organisms. Further, there is evidence that a highly specific serine carboxypeptidase may be involved in the regulation of blood pressure in humans by catalyzing removal of the C-terminal residue of angiotensin II and III [Odya et al., 1978; reviewed by Odya and Erdös (1981)].

These enzymes are of commercial interest as well. They can be used for C-terminal amino acid sequencing and for de novo protein synthesis, since they will catalyze transacylation reactions to nucleophiles other than water.

In order to provide a structural basis for the explanation of the many unusual properties of this class of enzyme, we have undertaken an extensive investigation of the structure–function relationships of two members of this class, monomeric yeast serine carboxypeptidase (CPD-Y, M_r 60K) and homodimeric wheat serine carboxypeptidase II (CPDW-II, M_r 120K). Here, we report on the refined structure of CPDW-II at 2.2-Å resolution and the results of a study of the binding of a reaction product arginine to the enzyme.

MATERIALS AND METHODS

Crystallization. Wheat serine carboxypeptidase II (CPDW-II) was purified as previously described (Breddam et al., 1987; Wilson et al., 1990). Large single crystals were obtained by the hanging drop vapor diffusion method from polyethylene glycol/sodium chloride solutions as previously reported (Liao & Remington, 1990). After being harvested, the crystals were equilibrated in a storage solution (0.5 M sodium chloride, 50 mM sodium acetate, pH 5.7). Polyethylene glycol is not required for crystal stability. If not used immediately, the crystals were stored at 4 °C.

Heavy Atom Derivatives and Data Collection. Potential heavy atom derivatives prepared by soaking crystals in modified storage solution were screened using precession photographs. Three native data sets and six derivative data sets have been collected at a variety of installations, include multiwire detectors at the University of California, San Diego (UCSD), Nicolet-Xentronics area detectors at the National Cancer Institute in Frederick Maryland (NCI), oscillation cameras at the University of Oregon, and a Fast area detector at the Brookhaven National Laboratory (BNL). Area detector data were reduced using the supplied UCSD data processing software (Howard et al., 1985), Xengen (Howard et al., 1987), and the MADNES package using three-dimensional profile fitting (Messerschmidt & Pflugrath, 1987; Kabsch, 1988). Oscillation film data were processed as described by Schmidt et al. (1981). Conventional X-ray sources were graphite monochromated Rigaku RU-200 rotating anode generators, permitting data to be collected to a maximum of about 2.7-Å resolution. At Brookhaven, the intensity of the synchrotron X-ray source made the collection of data to 2.2 Å feasible at a wavelength of 1.1 Å. All of the data were collected at room temperature except for the native synchrotron data set, which was at 4 °C. In order to determine whether data collected from different facilities were in adequate agreement, a native data set was tekan from each facility.

Heavy atom data sets were limited to 3.5-Å resolution to minimize data collection time and crystal deterioration. The soaking conditions that resulted in useful derivatives have been described (Liao & Remington, 1990).

Phase Determination. The heavy atoms positions were determined and refined using the PROTEIN package (Steigemann, 1974). The two major sites of chloro(2,2′:6′,2″-terpyridine)platinum(II) derivative (H1PT) were found by the automated Patterson search for self-consistent Harker and cross vectors in the difference Patterson map using the VECVER routine. The minor binding sites of H1PT and binding sites of other derivatives were determined by the usual difference Fourier method. The handedness of the space group was identified by comparing the observed and calculated anomalous differences for each derivative in the two possible enantiomorphic configurations. The heavy atom parameters were refined and a native best Fourier map was calculated at 3.5 Å.

New Solvent Leveling Procedure. We originally anticipated, on the basis of an assumed solvent content of about 50%, that the asymmetric unit would contain the dimer of CPDW-II. However, the MIR[1] map revealed a much larger solvent content, so the density was modified using a reciprocal space version of B. C. Wang's solvent flattening procedure (Leslie, 1987; Wang, 1985; S. L. Roderick, unpublished program), which yielded a greatly improved density map. Approximately 50 cycles of solvent flattening were performed. At the end of each cycle, phases from the transformed solvent leveled map were transferred to observed structure amplitudes.

We found that a minor modification of the usual procedure resulted in a better map, as judged subjectively. In each pass, after the solvent envelope is determined, the usual procedure is to set the solvent region to a constant value (often 0.0). Taking into consideration that the mask envelope may initially contain large errors, it seemed reasonable that the density in the presumed solvent region be treated more gently. In this procedure, the value of the electron density at a presumed solvent grid point was replaced by a fraction $S < 1.0$ (the "squeeze fraction") of the density at that grid point. S was adjusted downward toward 0.0 after each round of 10 cycles, in steps of 0.2 beginning at 0.8. As judged by rms phase difference between the final calculated and the original MIR phases (about 60°), the procedure converged. The best combination of the variable parameters (the squeeze fraction and assumed solvent content) were selected by inspecting skeletonized electron density maps. The clarity of the skeletonized representation of the helical regions was a very sensitive indicator of the quality of the map.

Model Building. A skeletonized version of the electron density map produced by the BONES program (Jones & Thirup, 1986) was inspected. Several α-helices of the correct hand were immediately recognized. The initial model was constructed on a computer graphics system on the basis of

[1] Abbreviations: MIR, multiple isomorphous replacement; rms, root mean square; MAN, mannose; FUC, fucose; NAG, *N*-acetylglucosamine.

both contoured and skeletonized density representations. The approach was to begin at helical regions and to guess the identity of the amino acid side chains. A sequence comparison program (Remington et al., 1982) was used to compare the X-ray sequence with the published sequence (Breddam et al., 1987). A 3.5-Å resolution, it is not possible to recognize peptide-plane orientations in turn and loop regions, so these were assigned on the basis of the α-carbon coordinates, using the DGLP fragment retrieval option of the FRODO program.

Crystallographic Refinement. The restrained least-squares reciprocal-space refinement package TNT (Tronrud et al., 1987) was used to refine the model. The refinement started with 3.5-Å resolution data and later was extended to 2.7- and 2.2-Å resolution as data became available. Large features in $F_o - F_c$ maps of 2.7- and 2.2-Å resolution refinement were inspected for model errors or as possible locations for bound solvent, while $2F_o - F_c$ maps based on model phases were examined to assess the reliability of the model. After each round of model building, 20–30 cycles of automated TNT refinement were performed.

Addition of Carbohydrate to the Model. Carbohydrate residues were first included in the model at the 2.7-Å resolution refinement, although electron density for these residues was apparent in the 3.5-Å solvent-leveled map. Models for the carbohydrate linkages were initially based on the haemaglutinin structure (Wilson et al., 1981; Brookhaven Data Bank code 1HMG). The model was constructed using theoretically derived torsion angle maps of the various glycosidic linkages (Bock, 1983) as guidelines.

Addition of Solvent Molecules. During refinement at 2.7-Å resolution, presumed water molecules were included in the model where difference electron density greater than two standard deviations was apparent, provided at least two hydrogen bonds could be made to the protein or other water molecules. Thirty water molecules were included for the refinement at 2.7-Å resolution. At 2.2-Å resolution, all of the peaks higher than four standard deviations in the $F_o - F_c$ map were located, and additional water molecules were included, provided at least one hydrogen bond could be formed.

RESULTS

Data Collection. CPDW-II crystallizes in space group $P4_12_12$ with $a = 98.6$ and $c = 210.0$ Å, with one monomer in the asymmetric unit, resulting in a packing parameter $V_m = 4.2$ Å^3 per dalton. Using laboratory X-ray sources, the native crystals of CPDW-II diffract to 2.3 Å and have a lifetime of five to seven days in the X-ray beam. However, the data between 2.7 and 2.3 Å are weak and decay rapidly. At Brookhaven, crystals diffracted to 2.2-Å resolution for about 6 h, permitting a reasonably complete data set to be collected from two crystals (Figure 1). The data statistics for the native as well as the four derivative data sets are listed in Table I. The isomorphous R factors shown in this table were calculated against the UCSD native data set. Table II shows the R factor between native data sets collected from different facilities at 3.5-Å resolution. Data sets collected at various laboratories agree with one another reasonably well, but surprisingly, the best agreement was between the University of Oregon film data and the UCSD area detector data.

Heavy Atom Parameters and MIR Statistics. The anomalous correlation values for the two choices of enantiomer are listed in Table III. The refined heavy atom parameters are available as Supplementary Material. The MIR phasing statistics (Table IV) shows that H1PT, PEDD, and HGCL

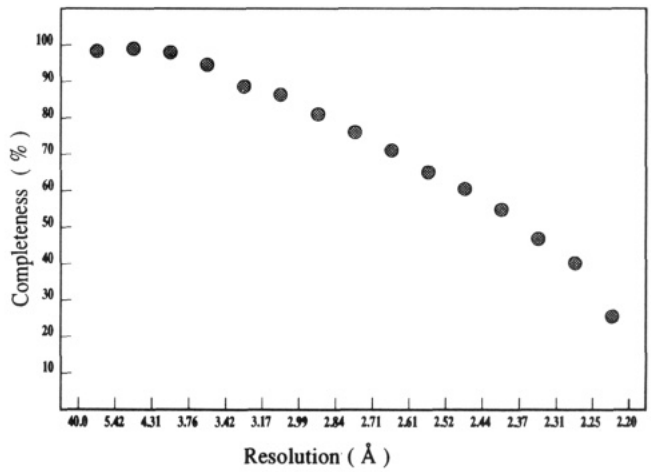


FIGURE 1: Completeness of reflection data (ratio of observed reflections to theoretically possible, expressed as a percentage) in 15 shells of resolution for the synchrotron data set collected at Brookhaven National Laboratory, Beamline X-12C.

are derivatives of moderate quality while DCL4 is a rather poor derivative.

Solvent Flattening. The initial 3.5-Å electron density map was obtained using MIR phases and had an overall figure of merit of 0.50. Figure 2a shows several superimposed sections of this map, representing a 5-Å slice in the Z direction and a full unit cell in X and Y. Although the general location of the protein molecule in the unit cell was clear, as were some features of secondary structure, the map would have been difficult to interpret (for example, the handedness of helical regions could not be discerned). However, after modification by the solvent flattening procedure, which was optimal assuming 70% solvent content (the actual content is 73%), the map was greatly improved, and several helical regions with the correct handedness could be seen (Figure 2b).

Initial Model and Refinement. The initial model consisted of 356 residues out of the 423 total. Owing to a lack of interpretable density, the side chains of 41 residues of the initial model were replaced by shorter ones, and no carbohydrate residues were included in the model. The conventional crystallographic R factor of this model was 42% of 10 484 reflections between 6.0- and 3.5-Å resolution. The course of refinement is listed in Table V. No restraints were imposed on torsion angles and chiral centers. Large errors in these two types of parameters are reported by the TNT program and were a useful indicator of serious errors requiring manual correction. However, convergence was exceptionally rapid, and few errors were detected in peptide plane orientation, which we attribute to the effectiveness of the DGLP procedure. The final refined model has a conventional R factor of 16.9% of all 35 941 reflections observed between 6.0- and 2.2-Å resolution. The root mean square deviation from ideal bond lengths is 0.018 Å and from ideal bond angles is 2.8°. The model contains 406 amino acid residues, 10 carbohydrate residues, and 199 water molecules for a total of 3533 non-hydrogen atoms. The average thermal parameter for main chain atoms is 28 Å^2 and for all atoms is 31 Å^2. There are three *cis*-proline residues in the final refined model, residues 43, 54, and 96. Disulfides are formed between residues 56–303, 210–222, and 246–268. The 17 amino acid residues missing from the model are at the C- and N-terminal regions of the two chains. In the final $2F_o - F_c$ maps, density is not apparent for these residues and a number of side chains. Some of these side chains are modeled with zero occupancy. Regions of doubtful accuracy are available as Supplementary Material.

Table I: Data Collection Statistics

compound	no. of crystals	detector[a]	resolution (Å)	no. of observations	unique reflections	R_{merge}[b]	R_{iso}[c]	completeness
NAT1	1	UCSD	3.5	90 719	13 423	4.3%		97.6%
NAT2	3	OSC	2.7	68 465	28 060	7.8%		84.3%
NATS	3	BNL	2.2	103 780	38 081	6.8%		72.9%
DCL4	2	UCSD	3.5	75 002	13 347	5.2%	9.0%	98.8%
H1PT	3	OSC	3.5	36 003	13 090	10.2%	14.1%	92.5%
HGCL	1	NCI	3.5	68 046	12 934	9.5%	11.4%	90.1%
PEDD	1	NCI	3.5	50 912	13 567	9.6%	14.0%	85.6%

[a] UCSD, multiwire area detector; OSC, oscillation camera at the University of Oregon; NCI, Nicolet-Xentronics area detector; BNL, synchroton light source, wavelength = 1.1 Å. [b] $R_{merge} = \sum_{h,i}|I_{h,i} - \langle I_h\rangle|/\sum_{h,i}I_{h,i}$, where $\langle I_h\rangle$ is the average intensity of the i observations of reflection h. [c] R_{iso} is the mean fractional difference, expressed as a percentage, of reflection intensities in the derivative data set versus the native data set.

Table II: Comparison of Native Data Sets Collected at Different Facilities

	OSC	BNL	NCI
UCSD[a]	5.4%[b]	6.9	6.9
NCI	7.7	9.0	
BNL	8.5		

[a] Data sets are UCSD, multiwire area detector; OSC, oscillation cameras at the University of Oregon; NCI, Nicolet-Xenotronics area detector; BNL, synchrotron light source, wavelength = 1.1 Å; Enraf-Nonius FAST area detector system. [b] The standard crystallographic reliability factor (R factor) measured on intensities expressed as a percentage, $R = \sum_{hkl}|I_1 - I_2|/(1/2)\sum_{hkl}(I_1 + I_2)$. There are approximately 12 000 overlaps between each data set at 3.5-Å resolution.

Table III: Anomalous Correlation Coefficient for the Two Possible Space Groups[a]

space group	H1PT	DCL4	HGCL	PEDD
$P4_12_12$	0.414	0.338	0.384	0.358
$P4_32_12$	0.020	0.001	0.049	0.031

[a] The anomalous correlation coefficient is as follows, where h are the reflection indices, and o and c refer to observed and calculated quantities:

$$C = \frac{\sum[F(h) - F(-h)]_o[F(h) - F(-h)]_c}{\sqrt{\sum[F(h) - F(-h)]_o^2\sum[F(h) - F(-h)]_c^2}}$$

Coordinates have been deposited in the Brookhaven Data Bank (code 3SC2).

Fold of the Polypeptide Backbone. The fold of the polypeptide backbone is as reported in our earlier 3.5-Å resolution study (Liao & Remington, 1990). It is an $\alpha+\beta$ protein, with 15 helices on either side of a mixed β-sheet consisting of 11 strands, of which the central six are parallel. Residues 180–310 appear to be an insertion of largely helical segments into the basic α/β topology and surround the active site which is located at the C-terminal end of strand 6. The higher resolution refinement revealed no serious errors in the topology reported for the 3.5-Å model. However, one error was discovered at residues 134–144, a central strand of the β-sheet. In this region, the chain was traced one residue out of register, which resulted in the incorrectly reported difference between the shape of the electron density and the amino acid sequence at residue 141 (Liao & Remington, 1990). A stereodiagram of the α-carbon backbone of the final refined model is shown in Figure 3. The Ramachandran diagram (Ramachandran & Sasisekharan, 1968) for main-chain torsion angles (Figure 4) shows few residues substantially outside allowable regions for non-glycine residues. Two of these, 308B at (ϕ,ψ) angles of (–105,–105) and 375 at (65,–130), are found in poorly defined surface loops where electron density is low and so are probably in error. However, all other outliers are found in the active site in regions of well-defined electron density. These are Ser146, a key catalytic residue, Gly52, Gly53, and Glu65. Ser146 is forced to take on (ϕ,ψ) angles of (40,–100) by the formation of the very tight β-turn (residues 145–147) at the end of strand 6 and start of the helix 147–163. The side chain of Ser146 is also in an unfavorable conformation, probably due to steric interference with the backbone and His397 (Figure 5). Residues 52, 53, and 65 may have crucial catalytic roles and probably contact substrate during the catalytic cycle, so these deformations are likely to be important.

The secondary structure was assigned on the basis of main-chain hydrogen-bonding pattern. The assignment, together with the sequence of CPDW-II, is shown in Figure 6. The numbering of the CPDW-II residues was based on a tentative alignment with CPD-Y, which has been more fully characterized biochemically.

Carbohydrate Model. CPDW-II has six N-linked glycosylation sites per subunit. Complex carbohydrate is attached to Asn105, Asn113, Asn247, Asn291, Asn297, and Asn426. The type of carbohydrate residues and the linkages between them have not yet been chemically analyzed. However, the core structure of this type of glycosylation site is highly conserved in plant glycoproteins, often including fucose residues (Lennarz, 1980). Carbohydrate residues were first included in the model at refinement cycle six. All of the glycosylation sites show density except the Asn426 site, which is in a flexible C-terminal region not included in the model. A fucose residue was included at both Asn105 and Asn113 sites, as suggested by the electron density. It appears to be bonded to O3 of the first *N*-acetylglucosamine (NAG) residue in each case as suggested by Lennarz (1980). The electron density and the model of the best resolved glycosylation site (at Asn106) are shown in Figure 7. The density at these sites is readily apparent at lower resolution (3.5 Å), but at higher resolution the density was much weaker than that of well-defined regions of the protein molecule, presumably due either to high mobility or heterogeneity. In the eleventh refinement cycle, models of these sugar residues were extensively rebuilt, and two glycosyl groups at Asn247 and Asn297 were deleted from the model because of weak density. The density at these two sites suggested that these groups have multiple conformations. The type of sugars and the linkages between them that appear to be consistent with the final electron density map and are included in the final model are summarized in Figure 7c. All of the carbohydrate residues are exposed to solvent and have large temperature factors (50 Å^2 or greater) except the first residue at Asn291 (Figure 7).

Crystal Contacts and the Dimer Interface. CPDW-II is a homodimer in solution, although apparently no reports have appeared regarding the role or importance of dimerization. Each molecule makes three crystal contacts with symmetry-related molecules in the unit cell. One of the crystal contacts, in which there are extensive interactions between the two

Table IV: Heavy Atom Phasing Statistics in Eight Resolution Shells[a]

resolution (Å)	15.41	10.37	7.81	6.27	5.23	4.49	3.93	3.50
Figure of merit	0.73	0.74	0.75	0.72	0.60	0.51	0.43	0.33
derivative								
H1PT phasing power	0.83	0.91	1.20	1.48	1.59	1.52	1.27	0.96
H1PT R_c	0.78	0.65	0.67	0.64	0.67	0.72	0.77	0.78
DCL4 phasing power	0.37	0.46	0.60	0.74	0.68	0.61	0.54	0.56
DCL4 R_c	0.93	0.91	0.94	0.94	0.94	0.93	0.92	0.92
HGCL phasing power	1.05	1.05	1.28	1.41	1.38	0.99	0.83	0.83
HGCL R_c	0.58	0.66	0.63	0.77	0.79	0.88	0.97	0.84
PEDD phasing power	0.81	1.02	1.36	1.51	1.07	0.73	0.70	0.73
PEDD R_c	0.76	0.75	0.66	0.66	0.82	0.87	0.89	0.91

[a] R_c is the standard centric Cullis R factor, phasing power is the rms heavy atom scattering divided by the rms lack of closure of the phase triangles. The overall figure of merit was 0.50.

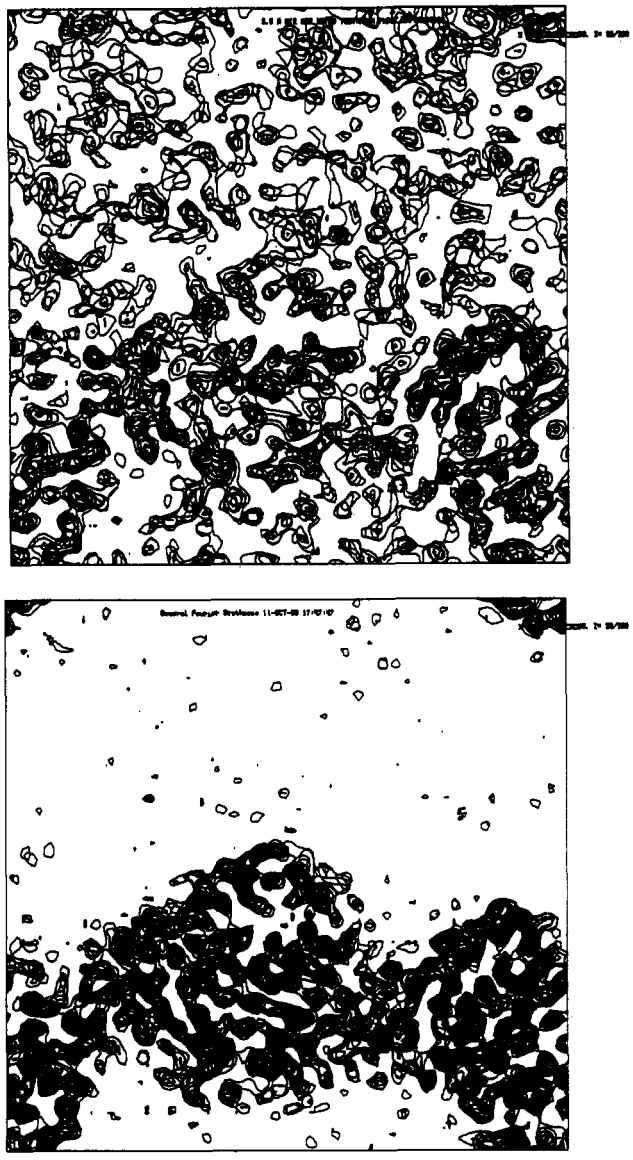


FIGURE 2: (a, top) Five sections (5 Å along Z) of the original multiple isomorphous replacement electron density map superimposed. The contours start at one standard deviation and are incremented by one standard deviation of the map. A full unit cell in X and Y is shown. Parts of two different molecules are seen in the lower part of the figure. (b, bottom) The same five sections of the electron density map after the solvent leveling procedure, as discussed in the text, was applied. The contour start and interval are one standard deviation.

molecules, must be the dimer interface of the enzyme. Most of the interactions in the interface are side-chain to side-chain, forming either direct hydrogen bonds between the two molecules or water-mediated hydrogen-bond networks. No

Table V: Course of Crystallographic Refinement

model no.	resolution range	no. of residues	B factor refinement	R factor	map inspected
0	6.0–3.5	352	initial model	42.0%	solvent leveled MIR
1	6.0–3.5	352	one overall	28.1	combined phase
2		396	one overall	21.1	combined phase
3		412	one overall	20.9	combined phase
4	6.0–2.7	412	one overall	25.8	$2F_o - F_c$ + 3.5-Å MIR
5		406	one/residue	21.5	$2F_o - F_c$ + 3.5-Å MIR
6		406	two/residue	21.0	$2F_o - F_c$ + 3.5-Å MIR
7		406	two/residue	19.7	$2F_o - F_c$ + 3.5-Å MIR
8	6.0–2.2	406	restrained	21.8	$2F_o - F_c$
9		406	restrained	19.9	$2F_o - F_c$
10		406	restrained	18.8	$2F_o - F_c$
11		406	restrained	18.5	$2F_o - F_c$
12		406	restrained	17.4	$2F_o - F_c$
13		406	restrained	17.5	$2F_o - F_c$
14		373	restrained	24.1	omit maps
15		406	restrained	17.0	$2F_o - F_c$
16		406	restrained	16.9	$2F_o - F_c$

interdigitation of the two monomers is observed. The interaction observed in each of the other two crystal contacts are water-mediated hydrogen-bond networks, and the contact regions are rather small. None of the glycosyl groups appear to be involved in any of the contacts.

Of the 15 753 Å^2 of accessible surface of the isolated monomer, slightly over 10%, or 1615 Å^2 are buried upon formation of the dimer. Atoms are defined as buried if their solvent accessibility changes by more than 5 Å^2 upon dimerization; this area is somewhat less distributed over nonpolar carbon atoms (756 Å^2) than polar nitrogen or oxygen atoms (810 Å^2). Surprisingly, the dimer interface is more hydrophilic than the rest of the surface of the molecule. Of the monomer surface not involved in dimerization, 7447 Å^2 are nonpolar carbon atoms, and 6614 Å^2 are nitrogen and oxygen. Perhaps this explains the rather low solubility of this enzyme, in water less than 0.1 mg/mL. At saturation in 0.5 M NaCl, 50 mM NaAc buffer, pH 4.0, the solubility is about 13 mg/mL.

The interaction between the monomers is illustrated by an α-carbon diagram (Figure 8) and is rather unusual. In each monomer, three helices (residues 190–203, 231–236, and 241–245) abut the surface of a β-sheet region in the other monomer (residues 336–337, 343–351, 368–382, 393, and 403), forming two interaction sites. The helix axes are approximately perpendicular to the sheet, quite unlike the usual packing, where the sheet strands are approximately parallel to the helix axes.

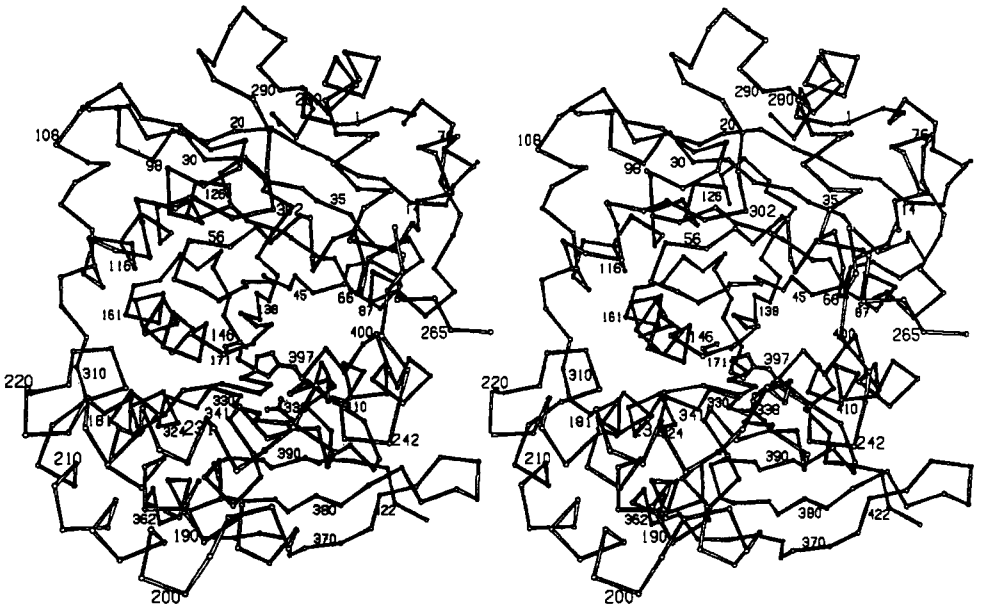


FIGURE 3: Stereo α-carbon trace of the backbone of CPDW-II. The point of view is into the active site. The active site residues Ser146, Asp338, and His397 are shown with filled bonds.

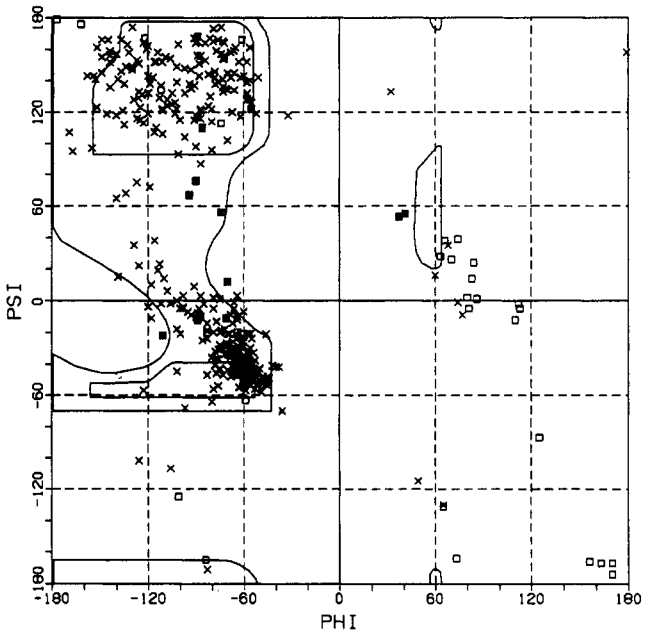


FIGURE 4: Ramachandran diagram of the final model of CPDW-II. Glycine residues are indicated with a box, asparagines are represented by a box and a cross, while other residues are represented by a cross. The "allowable" regions are contoured in the diagram. As discussed in the text, several non-glycine active site residues, including the active site serine, are in clearly defined but apparently strained conformations.

The dimeric model of the wheat enzyme may be useful in analyzing certain human genetic disorders, for example, galactosialidosis, which in several instances is known to result from mutations in the human lysozomal serine carboxypeptidase carbL, or "protective protein". Zhou et al. (1991) have shown that a mutation Phe412 → Val results in failure of the protective protein to dimerize, leading consequently to the incorrect assembly of a larger particle and loss of activity of β-galactosidase and neuraminidase. By sequence alignment, the residue in CPDW-II that corresponds to Phe412 is Trp380, a key residue centered in each interaction site (see Figure 8). The side chain lies in a hydrophobic cavity in the surface of the other monomer, making a weak (nonlinear) hydrogen bond with the carbonyl oxygen of residue 232. Substitution of this side chain for valine would introduce a cavity which would be likely to destabilize the dimer. The dimer of CPDW-II may therefore be a reasonable model for the dimer of the human enzyme.

Active Site. The two active sites in the dimer are located on either side of the dimer interface, face toward solvent, and do not appear to interact with each other. Each forms a large hemispherical pit, at the base of which the catalytic triad (Ser146–Asp338–His397) is found. The orientations of each side chain of the catalytic triad were clearly determined, with one exception: density for Oγ of Ser146 is weak and the temperature factor of this atom refined to over 50 Å^2, indicating some rotational freedom about the Cα–Cβ bond. The position of the Oγ atom should therefore be considered tentative. The main chain of this residue also is an unfavorable conformation although there is excellent density in this region (Figure 5).

The active site and the pattern of hydrogen bonds in this region are presented in Figure 9. Comparison of active site feature with those of trypsin and subtilisin leads us to suggest that an "oxyanion binding site" consists of residues 147 and 53, which are at the extreme C-terminal ends of strands 5 and 6 of the β-sheet. A solvent molecule, labeled W187, occupies this site and forms hydrogen bonds to the backbone amides of residues 147 and 53 in the unliganded structure.

Aspartate 338 forms hydrogen bonds to the amides of residues 340 and 341, the side chains of Asn176 and His397, as illustrated in Figure 9. The oxygen atom that is hydrogen bonded to the histidine accepts two other hydrogen bonds, from the amide of residue 341 and the side chain nitrogen of Asn176. Residue 176 appears to play a role similar to that of Ser214 of the trypsin family of enzymes [e.g., Warshel and Russell (1986), although McGrath et al. (1992) have called into question the importance of this residue]. However, in contrast to the trypsin or subtilisin families of enzymes, the plane of the carboxylate is far out of the plane of the imidazole ring. The His–Asp hydrogen bond makes an angle of approximately 45° to the plane of the carboxylate.

A major difference between the serine carboxypeptidases and the other serine proteinases is the solvent accessibility of the catalytic histidine residue (397). The side chain of tyrosine 239 makes an edge-on interaction with the imidazole and is in van der Waals contact. The total accessible surface area

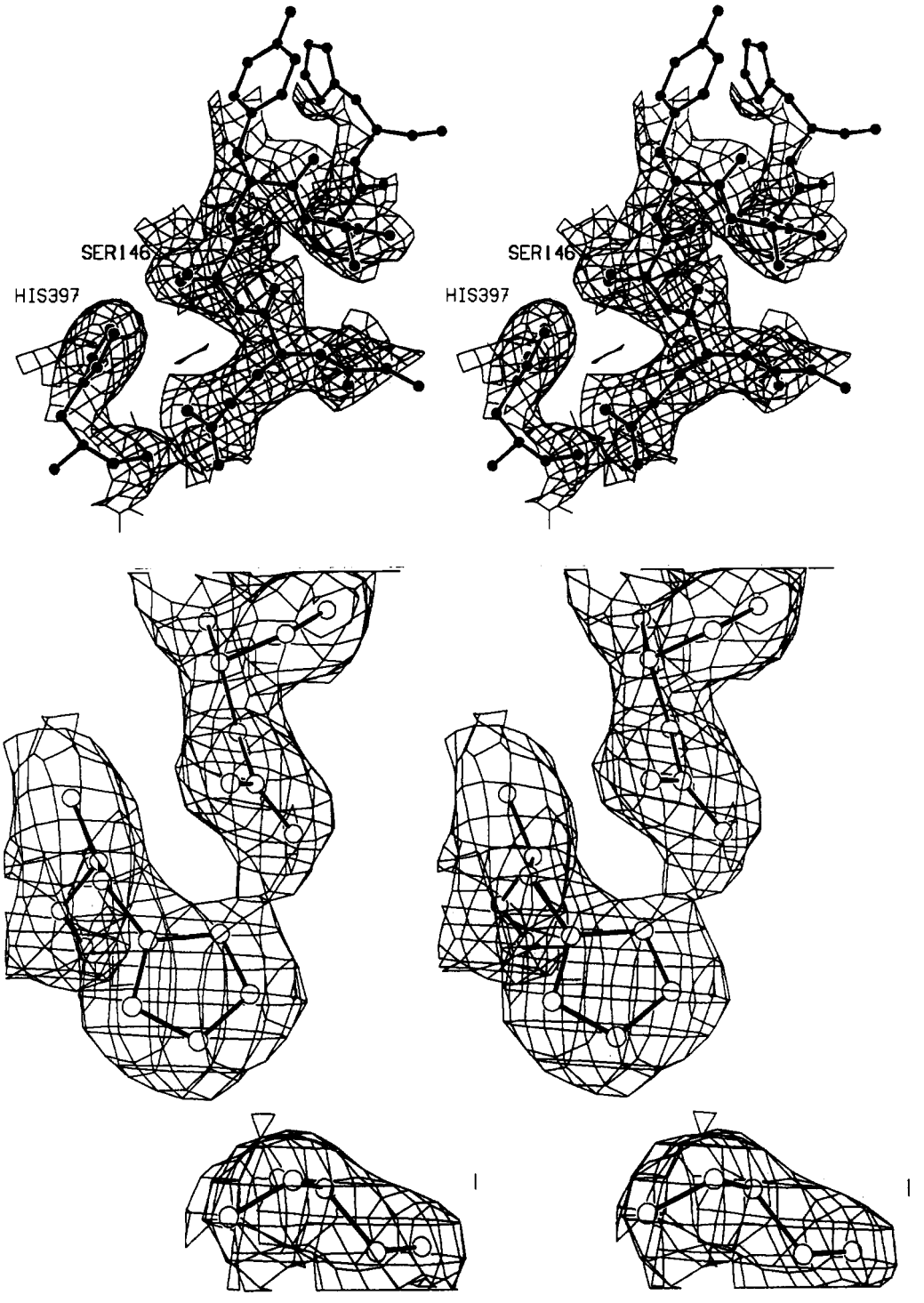


FIGURE 5: (a, top) Final $2F_o - F_c$ electron density map, contoured at one standard deviation, superimposed on the active site histidine and serine residues. The main-chain and side-chain conformational angles of Ser146 indicate an unfavorable conformation, but the electron density is well defined. The Oγ of the serine is to some extent disordered and has a large temperature factor. (b, bottom) As in panel a, but showing all the components of the catalytic triad Asp338–His397–Ser146 and their relationship.

of residue 397 is just 8 Å^2, with 5.3 Å^2 at Cδ2 and 2.7 Å^2 at Nϵ2. Model-building studies suggest that when substrate is bound, the histidine residue will be completely inaccessible to solvent.

The active site does not reveal exposed peptide carbonyl oxygens or amides which could interact with substrate by forming β-sheet hydrogen bonds, as do the trypsin and subtilisin families (Segal et al., 1981; Hirono et al., 1984; McPhalen & James, 1988). Instead, the cleft seems to be lined with predominantly hydrophobic side chains, suggesting that the mode of substrate recognition by serine carboxypeptidases is different than in subtilisin and trypsin.

Distribution of Charged Residues. Charged side chains usually are located on or close to the surface of globular proteins and are exposed to solvent (Miller et al., 1987; Barlow & Thornton, 1983). These residues are rarely found in the interior of proteins, except for those directly involved in catalysis. In the structure of CPDW-II, the majority of the charged residues other than the catalytic residue Asp338 are exposed to solvent on the surface of the molecule. However, a cluster of four glutamic acid residues is located near the active site in a region that is otherwise rather hydrophobic. The cluster consists Glu398, Glu64, Glu145, and Glu165. Glu145 and Glu65 form a hydrogen-bonding pair, as do Glu398

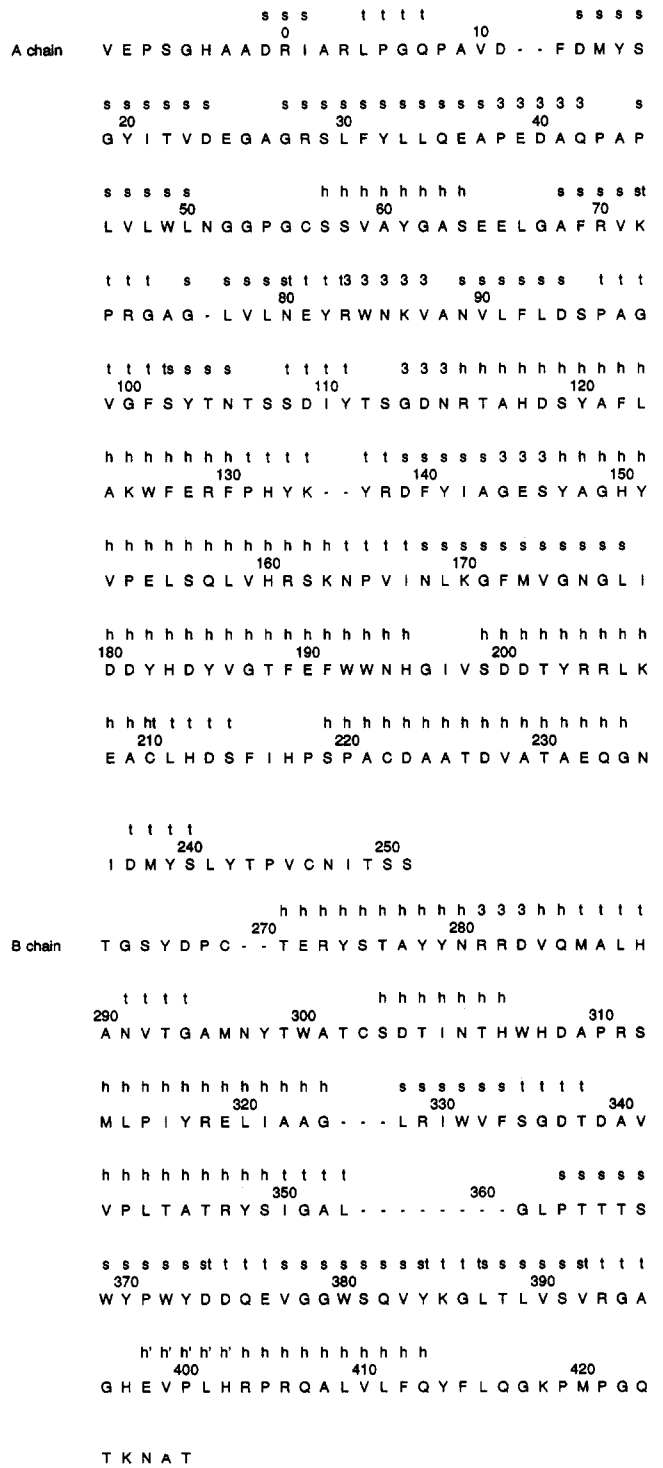


FIGURE 6: Secondary structure assignments and amino acid sequence (one-letter code) of CPDW-II. The symbols are s, β-sheet; h, α-helix; t, β-turn; and h', distorted α-helix. The criteria for assigning secondary structure is that the appropriate (ϕ,ψ) angles are achieved and at least one appropriate main-chain hydrogen bond is made. Residues which participate in two different types of secondary structure (in a transition) are indicated by assignments located between the amino acid codes. The residue numbers are those of the yeast serine carboxypeptidase, which is better characterized biochemically. Gaps indicate possible locations of insertions or deletions relative to the yeast protein and may be revised when the refined model of the yeast enzyme is available.

and Glu64 (Figure 10). There are several solvent molecules near the cluster, some with very low temperature factors, and one or more could possibly be a sodium or other cation. No positively charged residues appear near this region. It seems likely that this cluster of glutamic acid residues is involved in catalysis and/or stabilization of the active site configuration. In fact, Glu145 is adjacent to the active site Ser146, and its side chain is close in space to those of Ser146 and His397. In the next section we present evidence that Glu145 is involved in substrate binding, but additional roles are also possible.

Since these four glutamic acid residues are in an unusual environment that is only partially accessible to solvent and are clustered together, they may have unusually low pK_as. Under data collection conditions at pH 5.7, some or all of these side chains may be protonated and form hydrogen bonds with each other. Crystals are observed to be stable over a pH range of 5.0–10.0. Outside this range, the crystals dissolve immediately. At high pH, deprotonation and rearrangement of the active site acidic groups may initiate this behavior.

Arginine Binding Site. CPDW-II shows high preference for arginine and lysine at the P1′ position [notation of Schechter and Berger (1967) and Breddam et al. (1987)]. The binding of the free amino acid product arginine was investigated by difference Fourier methods, which demonstrated that a pocket formed by Tyr239, Tyr60, Glu398, and Glu272 forms the binding site for the side chain. Crystals were soaked in storage solution containing 100 mM arginine, pH 5.5, and 2.5-Å data were collected by oscillation photography. A total of 50 864 observations reduced to 24 182 independent observations with an R_{merge} of 0.074. Data were weak beyond 3-Å resolution, so an $(F_{o,Arg} - F_c)$ difference electron density map was calculated using model structure factors at 3.0-Å resolution. This map was not particularly clear, so the model, with all ligands removed from the active site, was refined aginst the $F_{o,Arg}$ data set to convergence. The new $(F_{o,Arg} - F_c)$ map was straightforward to interpret, so a model of the complex was constructed and refined. The final R factor is 0.192 for 19 029 reflections observed between 6- and 2.7-Å resolution with rms deviations of 0.014 Å from ideal bond lengths and 2.0° from ideal bond angles.

Arginine binds at the deepest pocket in the active site, with the carboxylate hydrogen bonded to Oϵ1 of Glu145 (2.4 Å), Nδ2 of Asn51 (2.9 and 2.5 Å), and the amide nitrogen of Gly52 (2.4 Å, see Figure 11). Interestingly, the amino group forms a hydrogen bond to His397 Nϵ2 (3.1 Å) as expected for the protonation of the leaving group during the reaction. The side chain stacks in the narrow channel formed by the aromatic side chains of Tyr239 and Tyr60 and appears to make generalized electrostatic interactions (i.e., does not directly form an ion pair) with the carboxylates of Glu398 (3.3 Å) and Glu272 (3.3 Å). This result must be interpreted with care, since a reaction product may not bind in the same manner as a productive substrate to the enzyme. However, the configuration immediately suggests an explanation for the pH dependence of the peptidase reaction (see Discussion).

Comparison of Serine Proteinase Active Sites. The catalytic role of the active site histidine as an acceptor of a proton from the serine Oγ is considered established. However, the role of the aspartic acid residue has been less clear. Blow et al. (1969) originally proposed a "charge relay" or double proton transfer mechanism, in which an additional proton was transferred between the histidine and the aspartate. This mechanism is now considered unlikely (Warshel et al., 1989; Bachovchin, 1986; Kossiakov & Spencer, 1981; Bachovchin & Roberts, 1978). The aspartate, which is essential (Carter & Wells, 1988, 1990; Sprang et al., 1987), is believed to play an electrostatic role, stabilizing the imidazolium cation developed during catalysis. The active site of CPDW-II is sufficiently different from those of trypsin and subtilisin (for example, if one attempts to superimpose all 24 main- and

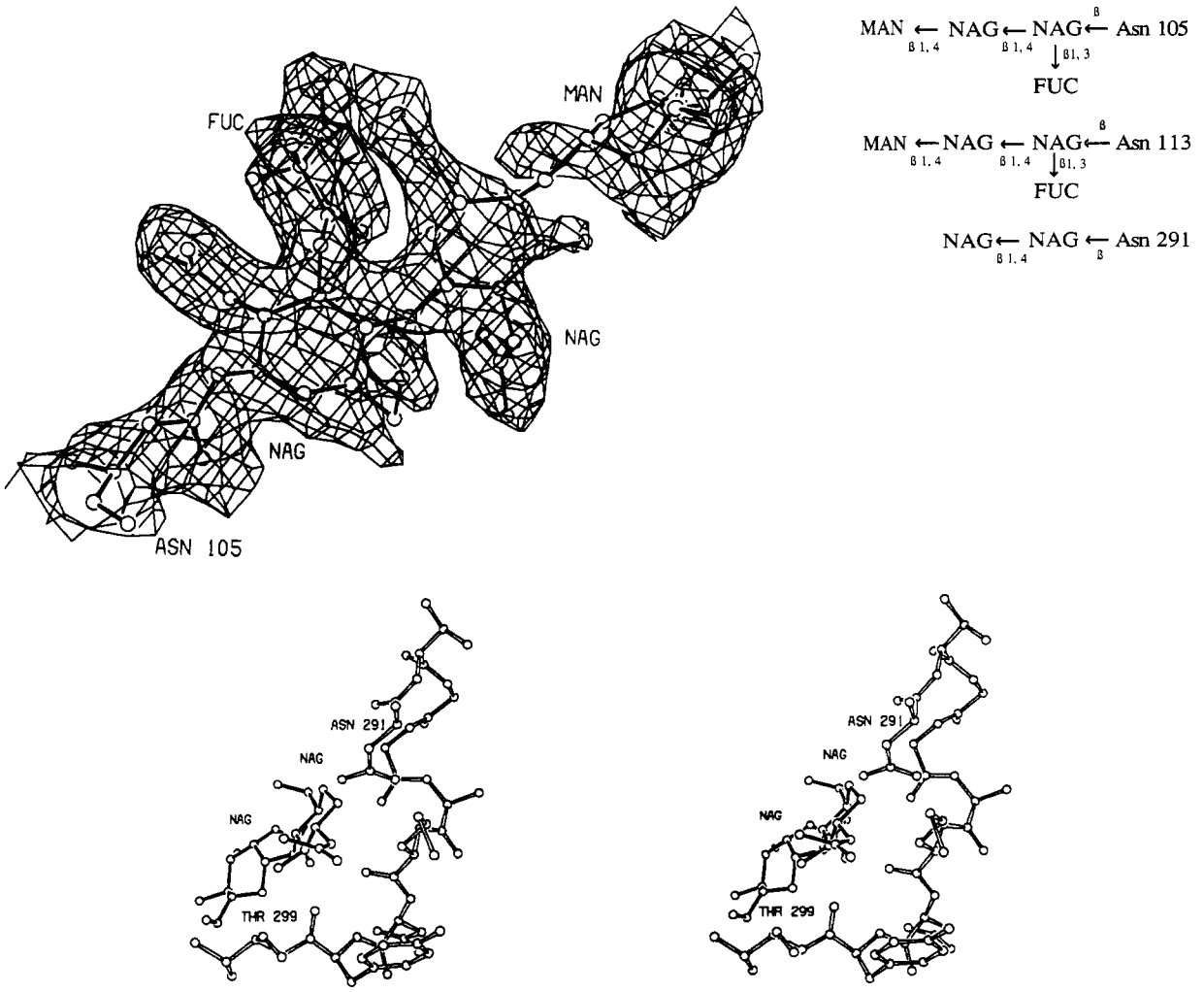


FIGURE 7: (a, top left) Electron density map contoured at one standard deviation superimposed on a proposed model of the glycosylation at Asn105. (b, bottom) Stereoview of the model of proposed glycosylation at Asn291. The first glycosyl group (filled bonds) appears to have very low solvent accessibility and is partially buried in the surface of the protein. There does not appear to be a fucose residue at the second position as proposed for other sites. (c, top right) Present model of the N-linked glycosylation, with linkage type and chemical composition indicated, at residues Asn105, Asn113, and Asn291. We emphasize that the model was derived from the electron density map alone and has not been confirmed by chemical analysis.

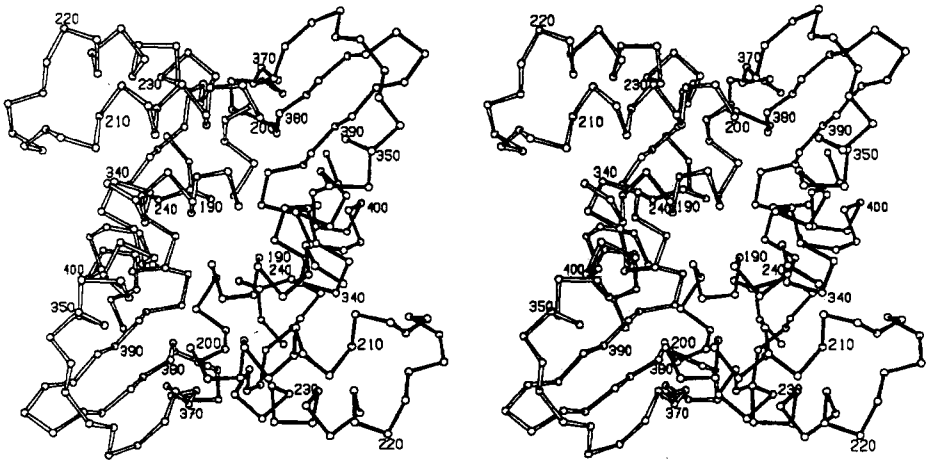


FIGURE 8: Stereo α-carbon trace of the dimer interface of CPDW-II, with one subunit indicated by filled bonds. There are two identical interacting surfaces, related by a crystallographic 2-fold axis, which consist of three helices of one monomer abutting three strands of β-sheet in the other monomer.

side-chain atoms of the catalytic triads of chymotrypsin and CPDW-II, the resulting coordinate error is 2.3 Å rms) that additional evidence in support for these more recent proposals is provided.

For the purposes of the comparison of these active site features, one high-resolution representative from each of the trypsin and subtilisin families was chosen. These were γ-chymotrypsin refined at 1.6-Å resolution (pH 7.0; Dixon &

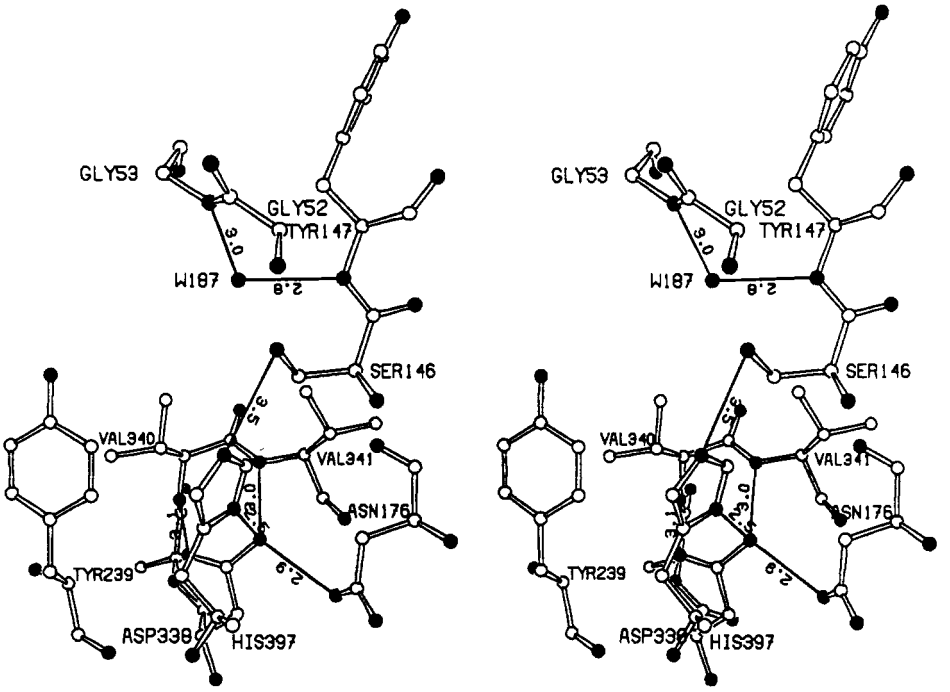


FIGURE 9: Stereodiagram of the hydrogen-bond network at the active site catalytic triad of CPDW-II. Hydrogen bonds have the indicated lengths in angstroms. A water molecule (W187), which occupies the putative oxyanion hole, is shown. Nitrogen and oxygen atoms are filled; carbon is open. The position of the serine Oγ and the indicated bond length of 3.5 Å to the histidine should be regarded as approximate due to disorder.

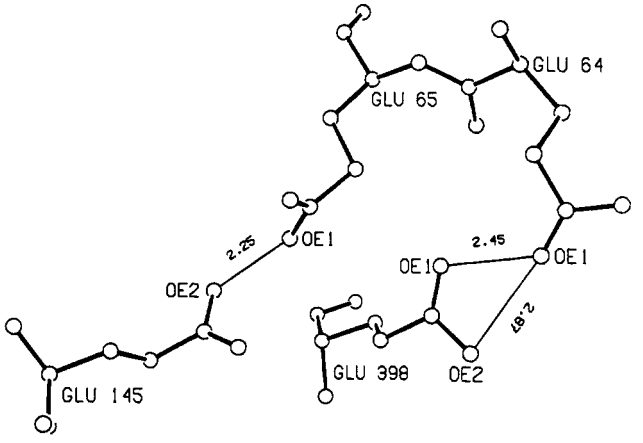


FIGURE 10: Apparent hydrogen-bond interactions between the four glutamic acid side chains in the active site of CPDW-II, indicating that at least two, and perhaps all, of these side chains are protonated at pH 5.7 (this study). Although the distance between Glu65 Oε1 and Glu145 Oε2 seems to be too short (2.25 Å), these groups appear to be well localized in the electron density map.

Matthews, 1989; Cohen et al., 1981), which contains a bound peptide, and Brookhaven Protein Data Bank entry ICSE, the complex of subtilisin Carlsberg with an inhibitor Eglin C from leech refined at 1.2-Å resolution (Bode et al., 1987). Selected atoms were superimposed by the method of Kabsch (1978).

Although the catalytic triads of subtilisin and chymotrypsin were originally thought to be similar (Kraut et al., 1972; Wright, 1972), the emergence of higher resolution refined structures reveal them to be more different than previously thought. McPhalen and James (1988) pointed out that, in subtilisin Carlsberg and α-chymotrypsin, the aspartates are in different positions with respect to the serine–histidine pair and that the hydrogen bonds between the aspartate and the histidine are not the same. However, in both structures, the carboxylate is very nearly coplanar with the imidazole ring, and the hydrogen bond to the Nδ1 position has nearly optimal geometry [see Table III of McPhalen and James (1988)].

In CPDW-II the His–Asp interaction is quite different. The plane of the carboxylate is far out of the plane of the imidazole (the Oδ2–Nδ1 virtual bond makes an angle of 45° with respect to the plane of the carboxylate), and the geometry is not optimal for proton transfer (Figure 9). However, the 2.55-Å Oδ2–Nδ1 distance does imply a strong electrostatic interaction, in support of the proposed stabilization of the imidazolium cation developed during catalysis (Bachovchin & Roberts, 1978).

On the other hand, the elements of the catalytic machinery which contact the substrate, namely, the serine Oγ, the histidine Nε2, and the nitrogen atoms forming the putative "oxyanion hole", superimpose remarkably well in the three enzymes. These six atoms (including the serine Cα and Cβ) superimpose with an rms coordinate error of about 0.4 Å in each case. Since this is roughly the combined coordinate errors of the three structural studies, these atoms can be considered to be in identical positions with regard to the substrate.

The catalytic groups were transformed according to this latter superposition, and the result is shown in Figure 12. From this figure, it appears that the aspartate can occupy virtually any position with respect to the histidine–serine pair and that the three structures differ mainly by a rotation of the Asp–His pair about the line connecting the histidine Nε2 and the serine Oγ. CPDW-II and subtilisin are most dissimilar, where the histidine ring plane and aspartate orientations differ by a 90° rotation about the Nε2–Oγ vector. We suggest that the aspartate–histidine pair and the histidine–serine pair should be regarded as catalytic diads that operate in concert. Indeed, in malate dehydrogenase, an Asp–His pair acts to deprotonate the hydroxyl of the substrate in an arrangement amazingly similar to that of the proteinases (Birktoft & Banaszak, 1982).

The significance of the strain in the active site of CPDW-II, i.e., the distorted backbone conformations of Glu65, Gly52, Gly53, and Ser146, is not clear. No evidence for strained

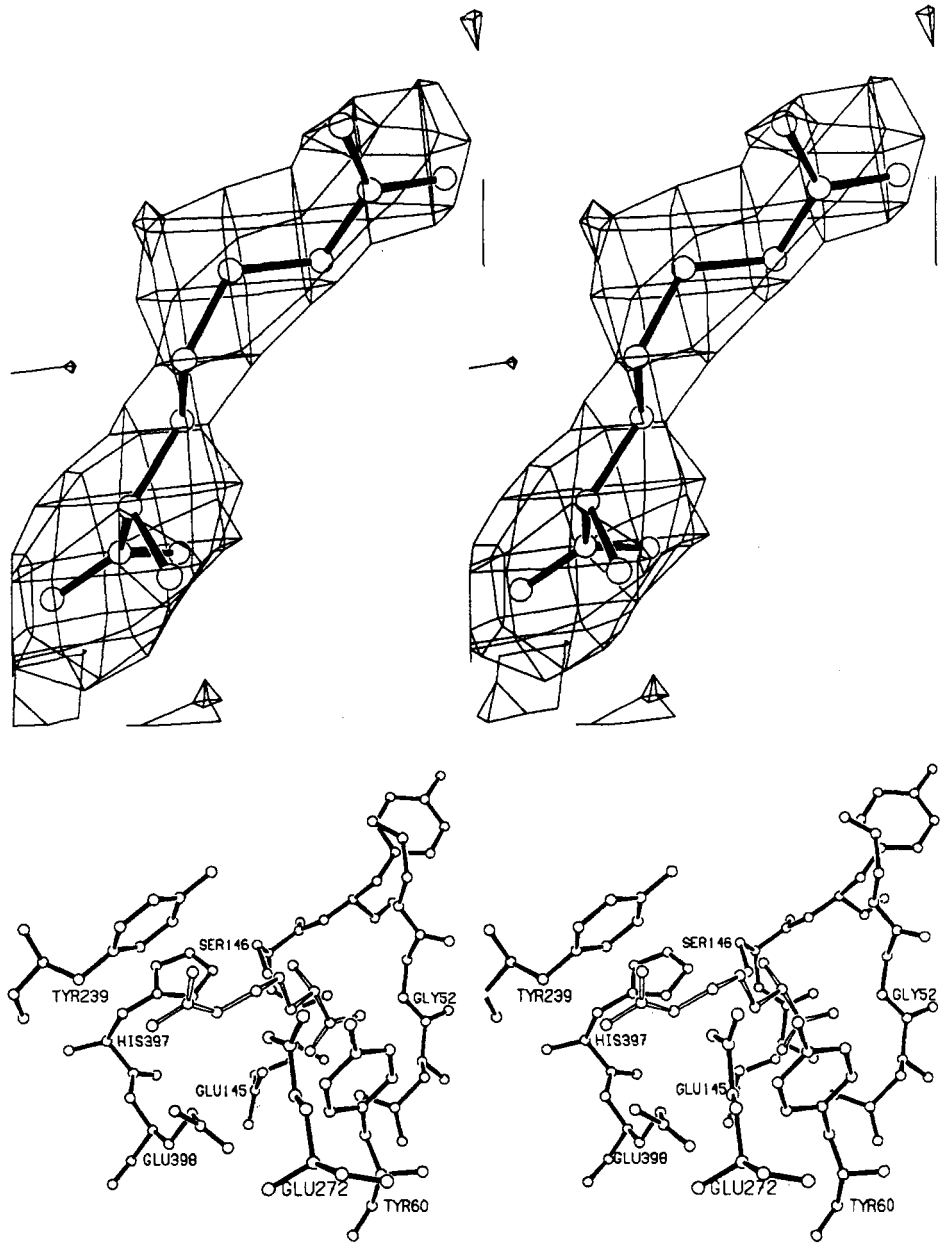


FIGURE 11: Stereoview of the leaving group arginine bound to the active site of CPDW-II. (a, top) Initial ($F_{o,Arg} - F_c$) difference electron density map contoured at two standard deviations after refinement of the protein model, not containing active site ligands, against the $F_{o,Arg}$ data set. (b, bottom) Active site residues are indicated with filled bonds, while the free amino acid is indicated with open bonds. The alkyl side chain stacks between Tyr60 and -239, while the guanidinium group interacts electrostatically with Glu398 and Glu272. The carboxylate of the arginine forms hydrogen bonds to the side chains of Glu145, Asn51, and the amide NH of Gly52, while the amino group forms a hydrogen bond with His397.

backbone conformations is present in the 1.6-Å resolution structure of γ-chymotrypsin, so this is not a requirement for serine proteinase activity. It is possible that the special constraints of binding and hydrolyzing a C-terminal residue, which has a functional group not found on substrates cleaved by endopeptidases, could require such deformations. Alternatively, the particular fold of this enzyme may require certain groups to be in high-energy configurations, developed at energetic cost during the protein folding process, to achieve the required catalytic efficiency.

DISCUSSION

Biochemical studies of serine carboxypeptidases [reviewed by Breddam (1986)] had indicated that a serine (Hayashi, 1973) and a histidine (Bech & Breddam, 1989) are essential for catalysis and suggested that this class of enzyme may employ the same catalytic mechanism as that of the much-studied serine endopeptidases. However, no evidence has been provided for an aspartic acid residue, a key residue in serine endopeptidases, in the active site of the enzyme. It was not clear that one would be required, as the optimum pH for peptidase activity of this class of enzyme is acidic, which is in contrast to the serine endopeptidases, for which the optimum pH is basic. Because of this, this group of enzymes was originally termed "acid carboxypeptidases" (Zuber & Matile, 1968). These points make it conceivable that serine carboxypeptidases may in fact have a different catalytic mechanism, despite the fact that a serine and a histidine are required for catalysis. These considerations provided part of the impetus for this structure determination.

The refined atomic model of CPDW-II at 2.2-Å resolution confirms the initial suggestion (Liao & Remington, 1990)

that a catalytic triad (Ser146, His397, Asp338), does exist in the protein. These residues are conserved in all serine carboxypeptidases of known sequence (Sorenson et al., 1987; Galjart et al., 1988). The structure also shows that these groups are in an arrangement similar to those of the serine endopeptidases, with the exceptions noted previously. Furthermore, an "oxyanion hole", which is thought to stabilize a negative charge on the tetrahedral intermediate in the catalytic reaction, appears in a similar position relative to the serine side chain. Thus it seems certain that serine carboxypeptidases utilize the same catalytic mechanism as that of serine endopeptidases. However, the specific roles of the key residues in question have been the subject of nearly continuous debate since the early structural studies (Blow et al., 1969), in particular the question of the protonation states of the histidyl and aspartyl side chains and how these may change during the course of the reaction [e.g., Bachovkin and Roberts (1978), Kossiakoff and Spencer (1981), Carter and Wells (1988, 1990), Sprang et al. (1987), and Warshel et al. (1989), to mention only a few]. Structural studies of a third, convergently related family of serine proteinases provide the opportunity for more extensive comparison of active site features than was previously possible. The refined atomic model of CPDW-II will serve as a basis for an ongoing investigation into the properties of this family of enzymes.

Evolution of Serine Carboxypeptidases. The trypsin and subtilisin families of serine proteinases were long ago recognized to be an example of convergent molecular evolution, since they have similar catalytic groups in the active site but have no sequence or structural homology. The order of the catalytic groups along the polypeptide chain differs as well. However, they share similar active site features and almost certainly employ the same catalytic mechanism. This provided the first strong evidence for convergent evolution at the molecular level, and several other examples have since followed. Clearly, serine carboxypeptidases are a third example of convergent evolution of serine proteinases, in that they have a completely different fold than either of the other families but have essentially the same active site features. Furthermore, it is now seems apparent that there is yet another family of serine proteinase, the acylpeptide hydrolase, or prolyl oligopeptidase family, which has also been shown to contain active site serine and histidine residues [e.g., Rawlings et al. (1991), Polgar (1991), and Jones et al. (1991)].

Recently, however, it was found that CPDW-II shares similar topology with four other enzymes with nearly unrelated activities. These are dienelactone hydrolase from *Pseudomonas sp. B13* (Pathak et al., 1988), haloalkane dehalogenase from *Xanthobacter autotrophicus* (Franken et al., 1991), acetylcholine esterase from *Torpedo californica* (Sussman et al., 1991), and lipase from *Geotrichum candidum* (Schrag et al., 1991). All of these enzymes have essentially the same fold of the polypeptide backbone, and a catalytic triad, but only the histidine is common to all three. There is no other sequence homology among these enzymes (except that the lipase and acetylcholine esterase do have limited sequence homology). The fold of this family of enzymes has been termed "the α/β hydrolase fold" since hydrolytic activity is the only common feature of the various enzymatic reactions (Ollis et al., 1992). Therefore, serine carboxypeptidases are also the product of a widely divergent course of evolution in which the same protein fold has acquired different catalytic activities. Thus these diverse protein structures offer a particularly rich insight into the flexibility of molecular evolution.

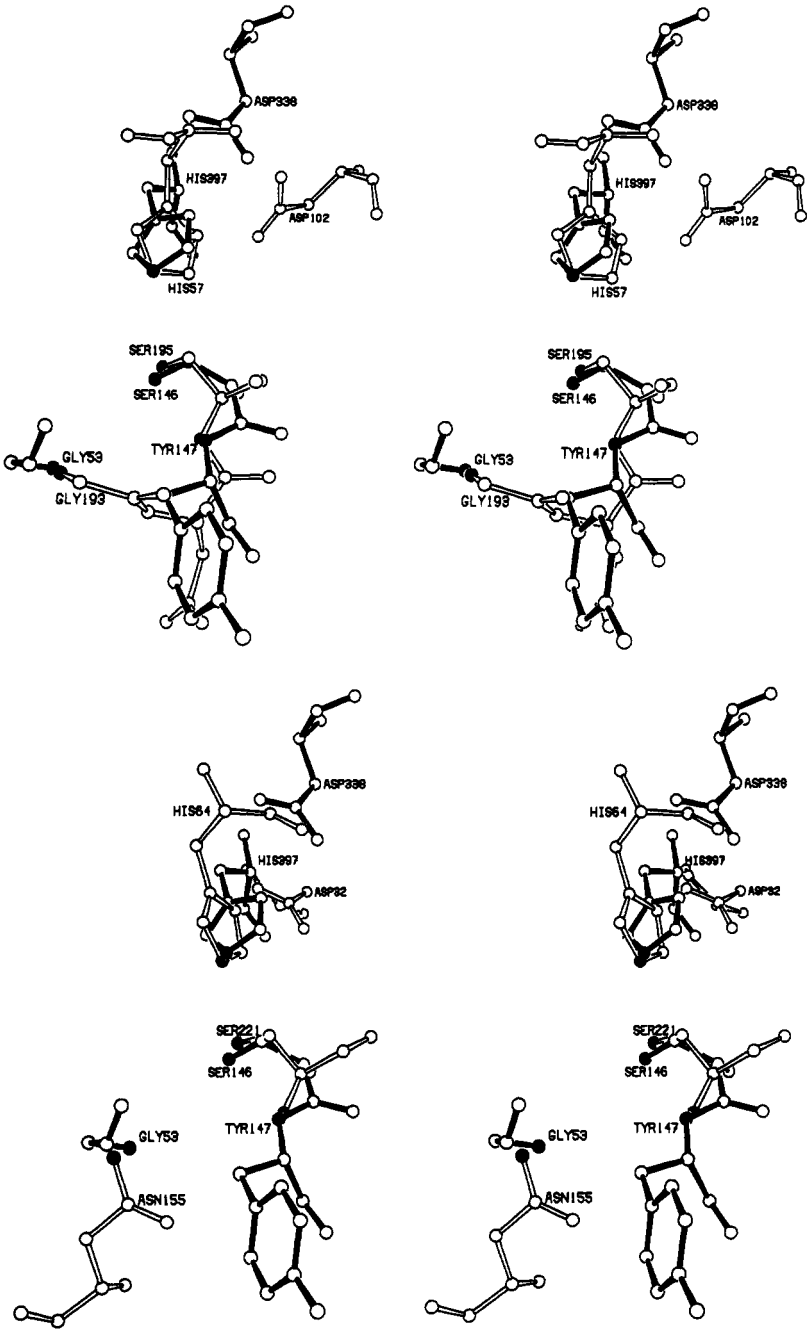
pH Dependence of Catalytic Activity. As pointed out in the introduction, serine carboxypeptidases have acidic peptidase activity with maximum activity at about pH 4.5–5.5, depending on the enzyme, and activity falls off at low pH consistent with titration of a group with a pK_a in this range. Indeed, there is measurable activity as low as pH 3.0 (Breddam, 1986), whereas trypsin is essentially inactive below pH 7.0.

If the titratable group is the active-site histidine, its pK_a is unusually low. Upon visual comparison of the active-site environment of the three families of serine proteinases, it becomes clear that in all three, the histidine side chain in the active site is not fully exposed to solvent and that the histidine of CPDW-II is less accessible to solvent (8.2 Å^2 total accessible area with only 2.7 Å^2 at Nϵ2) than that of either trypsin (70 Å^2 total, 6.3 Å^2 at Nϵ2) or subtilisin (36 Å^2, 4.0 Å^2 at Nϵ2). This additional shielding CPDW-II is due to tyrosine 239 (Figure 11), which is a conserved residue in serine carboxypeptidases of known sequences, which is also involved in substrate binding.

This lowered accessibility may acount for an *apparent* perturbation of pK_a of the active site histidine by differential partitioning of proton concentration between the aqueous phase and the relatively hydrophobic active site. However, it seems more likely that the pK_a of the histidine is perturbed by other factors, such as electronic effects from the partial charge distribution of aromatic ring of Tyr239 or other neighboring side chains. Detailed electrostatic calculations may reveal more about the pH dependence of the activity, but these must await the determination (by other techniques) of the charge state of the unusual cluster of acidic groups in the active site.

Kinetic studies of several serine carboxypeptidases suggested that a side chain with an apparent pK_a value of 5.5 functions as the anion-binding site for C-terminal carboxylate group of peptide substrates [reviewed by Breddam (1986)], possibly a histidine. However, there is only one histidine conserved in different serine carboxypeptidases, and it is a catalytic residue. The present model has no other positively charged side chain located near the active site; however, a pair of conserved glutamate residues, Glu145 and Glu65, which interact with each other, are close to the active site serine. One or both of these residues must therefore be uncharged. The side chain of Asn51, which is located in a very highly conserved section of the amino acid sequence, forms a number of hydrogen bonds and appears to be very rigidly located, with the Nδ2 located 3.4 Å from Glu65 Oϵ1. The orientation of the side chain of Asn51 is unambiguous, as Oδ1 accepts hydrogen bonds from the amide of residue 61 (2.7 Å) and the indole nitrogen of Trp49 (3.0 Å). The location and orientation of Glu145 and Asn51, which are conserved in serine carboxypeptidases, form the C-terminal carboxylate binding site. In the refined model of the enzyme, a solvent molecule (modeled as water) with a temperature factor of 42 Å^2 occupies this pocket and forms hydrogen bonds to the amide nitrogen of residue 52, Nδ2 as Asn51 and Oϵ1 of Glu145.

The structure of the enzyme with bound arginine demonstrated that this is a ligand-binding site and that the side chains of Glu145, Asn51, and the amide of Gly52 form hydrogen bonds with the C-terminal carboxylate. Therefore, it seems likely that the deprotonation of one of the glutamic acid residues, either Glu145 or Glu65, or both would disrupt the hydrogen-bond network and may be responsible for the increase of K_m for peptides at high pH, as a result of eitherthe negative charge or structural perturbations. The other pair of glutamic acid residues in the vicinity, Glu64 and Glu138, is not conserved in other serine carboxypeptidases and so is



unlikely to be directly responsible for the pH dependence of peptide binding.

Basis for Substrate Specificity. The structure of the arginine binding pocket is such that only a C-terminal amino acid amide or smaller leaving group can be accommodated. Tyr60 forms a blunt wall at the end of the P1′ binding site. Therefore, the lack of endopeptidase activity is simply due to the fact that larger substrates cannot be accommodated in this site. The specificity of CPDW-II toward peptide substrates is for large side chains, preferentially lysine and arginine at the P1′ leaving group position (Breddam et al., 1987). However, substrates with hydrophobic side chains like Met, Phe, and Ile in P1′ are also hydrolyzed, but at about 10% of the rate of Lys or Arg. The basis for this now seems clear. Tyr239 and Tyr60 form a narrow channel on the surface of the enzyme to accommodate the hydrophobic part of the side chain, while Glu398 and Glu272 form a negatively charged pocket at the end of the channel which can interact with the terminal charges of Lys or Arg but not the shorter side chains like Met and Ile. It is therefore not necessary to suggest that there are subsites within each side-chain binding site to accommodate the differing groups, and it seems likely that the mode of arginine binding will be representative of hydrophobic residues as well.

At the P1 position, ester substrates with Phe are preferred over by about a factor of 3 over those of Lys and Arg. The structure of the active site region does not immediately suggest an explanation for this observation, as there is no obvious "pocket" which would accommodate either sort of side chain as the P1′. There is an extended, rather broad hydrophobic depression which could accommodate at least two amino acids in P2 and P1 positions, suggesting that specificity may also include a P2 subsite. However, analysis of binding in this region of the active site must await the results of further inhibitor-binding studies, which have recently been initiated.

On the Significance of *G–X–S–X–S/A.* It has been noted that the serine proteinases have a pentapeptide "motif" with the sequence Gly–X–Ser–X–Gly/Ala containing the active

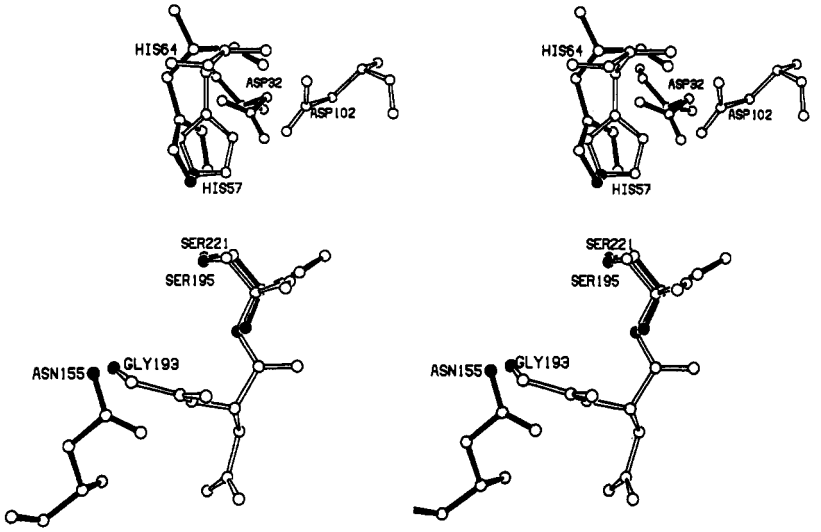


FIGURE 12: Three-way superposition of the critical active-site elements of representatives of the three families of serine proteinases. Shown are segments of γ-chymotrypsin, subtilisin Carlsberg, and CPDW-II. (a, opposite page top) Stereoview of the superposition of the "oxyanion holes" of chymotrypsin (open bonds) and CPDW-II (filled bonds). Active site residues of the catalytic triad are shown. The superposition was based on the two amides of the oxyanion hole, the serine Cα, Cβ, and Oγ, and the histidine Nε1. Key atoms in the histidine, serine, and oxyanion hole are filled. Note how the aspartate residues are placed differently by this superposition. (b, opposite page, bottom) As above, a superposition of the oxyanion holes of CPDW-II (filled bonds) and subtilisin (open bonds). (c, above) As above, superposition of chymotrypsin (open bonds) and subtilisin (filled bonds).

Table VI: Consequences of Alanine Substitution for Active Site Glycine Residues[a]

protein	residue substituted	bad contacts (Å)	to atom
chymotrypsin	Gly193 → Ala	2.48	Met192 O
		3.05	Gly142 CA
		3.17	Thr151 CG2
	Gly197 → Ala	1.72	W3 HOH
		2.88	Pro198 CD
		3.04	Asp194 O
CPDW-II	Gly144 → Ala	2.79	Glu145 N
		2.73	Leu50 O
		3.06	Val152 CG2
		3.07	Ala143 C
subtilisin	Gly219 → Ala	2.07	Leu145 O
		2.89	Asp146 CA
		2.46	Asn218 O
		2.77	Asn218 C
		2.99	Thr220 N
		2.96	Leu145 C

[a] Bad contacts made by alanine Cβ are listed.

site serine. The chymotrypsin-like enzymes have glycine in the fifth position, whereas subtilisins and serine carboxypeptidases have Ala in this position. A number of authors have suggested that proteins containing this motif may have peptidase or esterase activity [e.g., Brenner (1988)]. This suggestion implies that the glycine residues play a special role in the pentapeptide, possibly permitting an unusual conformation of the loop that is optimal for catalytic activity. Indeed, Brenner (1988) suggested that gene fragments coding for these peptides (and, by inference, catalytic activity) may have been transferred many times as a result of intron/exon shuffling and that these pentapeptides arose via divergent evolution. On the basis of the following observations, we suggest that the sequence similarities at positions 1 and 5 in these pentapeptides are fortuitous.

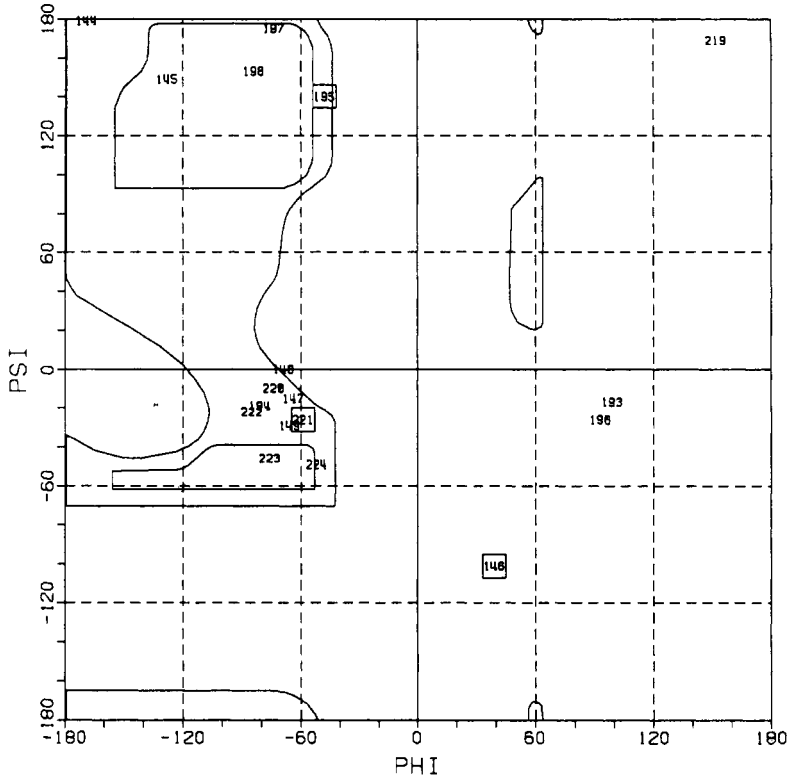
Examination of the three models on hand reveals that the glycine and/or alanine residues at positions 1 and 5 of the pentapeptides are very important. We replaced the glycine positions with alanine in a model-building experiment and observed severe steric clashes (Table VI); therefore, these residues must be small. Also, three of the glycines (193 and 196 in chymotrypsin and 219 in subtilisin) have conformational angles not allowed for other residues (Figure 13a). However, as is clear from Figure 13 and Table VII, the conformations of the three pentapeptides are not very similar. The serine residue is always the second residue of a β-turn, but the turn type is not conserved. In CPDW-II, the turn type is II′ (Lewis et al., 1972), while in chymotrypsin and subtilisin the turn type is II and I, respectively (Table VII). In fact, the conformational angles of the serines fall in completely different parts of the Ramachandran diagram (boxed residues in Figure 13a).

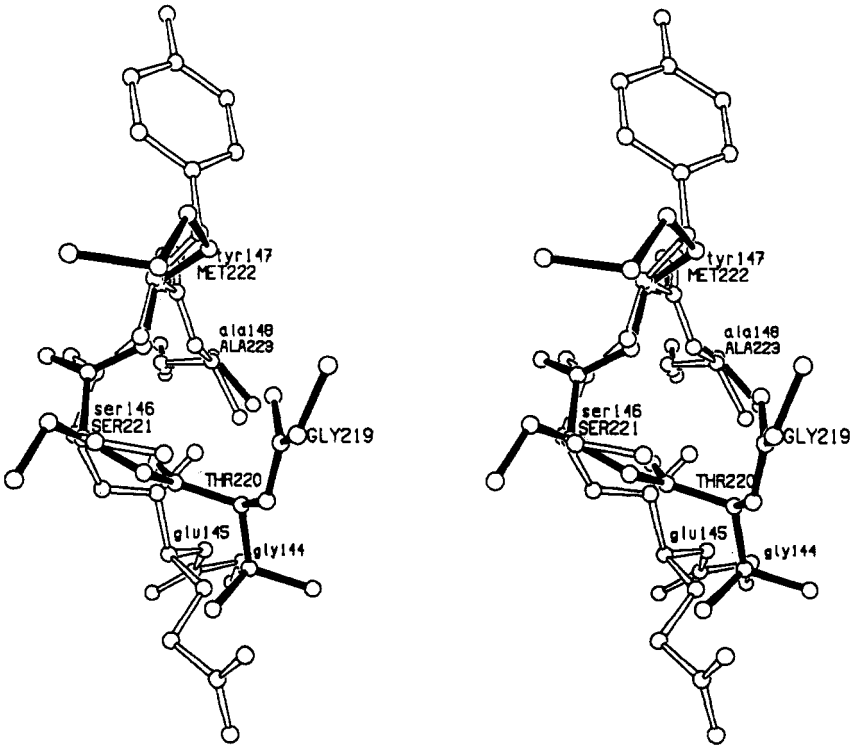
In more detail, the steric clashes upon glycine to alanine substitution would take place with different parts of the respective proteins and not always within the pentapeptide itself. This is illustrated by the superpositions in Figure 13b,c.

Glycine is required at position 5 in chymotrypsin for two reasons: the conformation angles of Gly196 are wrong for Ala, and severe steric clashes would result (Table VI) if this were larger. Why is alanine permitted in the other two enzymes? In Figure 13b, the pentapeptides of subtilisin and chymotrypsin are superimposed. The ψ angles of the serines differ by about 180°, placing residues 222 and 223 in subtilisin to a quite different location in which alanine (but not a larger residue) at position 5 can be accommodated. In Figure 13c, the same is seen to be true for CPDW-II. Thus, it seems that the ψ angle of the serine at position 3 dictates the glycine/alanine choice at position 5.

Table VII: Main-Chain Conformational Angles of Active Site Pentapeptide GXSXG/A

protein	φ1	ψ1	φ2	ψ2	φ3	ψ3	φ4	ψ4	φ5	ψ5	turn type
chymotrypsin 193–197	99	−17	−81	−19	−48	140	93	−26	−74	175	II
CPDW-II 144–148	−169	179	−128	149	39	−101	−64	−15	−69	0	II′
subtilisin 219–223	152	169	−74	−10	−59	−26	−85	−22	−76	−46	I





The conformations of the glycine residues at position 1 are quite different in the three proteins. In subtilisin and chymotrypsin (Figure 13b), the ψ angles of Gly193 and -219 differ by 180°, so the side chains of these two residues, if larger, would clash with completely different parts of the respective proteins (Table VI). Comparing CPDW-II and subtilisin, we find that the ϕ angles of the serines differ by about 90° (Figure 13c), so that the preceding residues are in very different positions. Nevertheless, alanines at either position would severely disrupt the active site.

To summarize, there are no structural features in common between these three different serine proteinases supporting the idea that the pentapeptides with the G–X–S–X–G/A motif diverged from a common ancestor. This should not be a surprise, since even the seeming requirement for an aspartate in the Asp–His–Ser catalytic triad has also recently been shown to be fortuitous, as acetylcholine esterase (Sussmann et al., 1991) and *G. candidum* lipase (Shrag et al., 1991) contain a Glu–His–Ser catalytic triad. We will need a much larger structural database in order to draw an accurate picture of the course of molecular evolution.

We have recently solved the structure of yeast serine carboxypeptidase (CPD-Y, or proteinase C in older literature) at 2.7-Å resolution by multiple isomorphous and molecular

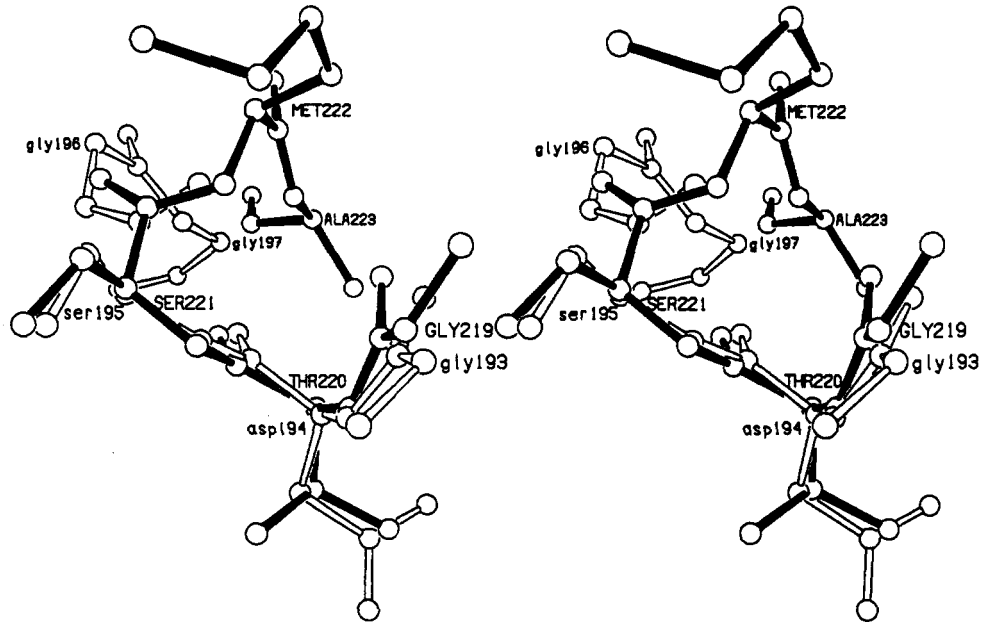


FIGURE 13: Structural evidence that the active site pentapeptides of chymotrypsin, subtilisin, and CPDW-II did not arise via divergent evolution, as suggested by the sequence motif G–X–S–X–G/A. (a, opposite page top) Ramachandran diagram of the pentapeptides containing the active site serine of chymotrypsin (residues 193–197), subtilisin (219–223), and CPDW-II (144–148). The (ϕ,ψ) values are centered on the plotted number, and the active site serines are boxed to emphasize their different conformations. (b, opposite page, bottom) Stereodrawing of a superposition of the active site pentapeptides of CPDW-II (open bonds) and subtilisin (filled bonds). The superposition was performed by hand to emphasize the similarity of the conformation of residues 146, 147, and 148 of CPDW-II with that of 221, 222, and 223 of subtilisin and the abrupt change of conformation proceding the Cα of the serines. (c, above) As above, stereodrawing of active site pentapeptides of subtilisin (residues 219–223) superimposed on chymotrypsin (residues 193–197). The drawing emphasizes the similarity of the conformations of residues 219–221 with those of 193–195, but there is an abrupt change of conformation due to a difference of 180° in the ψ angles of the two serines.

replacement. The model is being refined. The results of the study and a comparison of the structures of the two enzymes will be presented elsewhere. We intend to use the yeast proteinase as a system for generation of selected mutants to test some of the hypotheses presented here.

NOTE ADDED IN PROOF

After submission of this paper, our attention was drawn to a comparison of active-site features of the *Rhizomucor miehei* triacylglycerol lipase and the serine endopeptidases (Derewenda & Derewenda, 1991). The lipase is structurally similar to CPDW-II. Using similar lines of reasoning, the authors reached the same conclusion regarding the fortuitous nature of the occurrence of the sequence G–X–S–X–G/A in these various enzymes. The active site serine of the *R. miehei* lipase has the same unusual (ϕ,ψ) angles as the serine in CPDW-II and the nucleophiles of the other "α/β hydrolases".

ACKNOWLEDGMENT

Diffraction data for this study were collected at Brookhaven National Laboratory in the Biology Department single-crystal diffraction facility at beamline X12-C in the National Synchrotron Light Source. This facility is supported by the United States Department of Energy, Office of Health and Environmental Research. Grants from the National Science Foundation Instrumentation and Instrument Development program and several pharmaceutical companies have supplied computing equipment for the facility.

SUPPLEMENTARY MATERIAL AVAILABLE

Two tables describing heavy atom parameters and regions of doubtful coordinate accuracy in the refined atomic model (2 pages). Ordering information is given on any current masthead page.

REFERENCES

Bachovchin, W. W., & Roberts, J. D. (1978) *J. Am. Chem. Soc. 100*, 8041–8047.
Barlow, D. J., & Thornton, J. M. (1983) *J. Mol. Biol. 168*, 867–885.
Bech, L. M., & Breddam, K. (1989) *Carlsberg Res. Commun. 54*, 165–171.
Birtkoft, J. J., & Banaszak, L. J. (1982) *J. Biol. Chem. 258*, 472–482.
Blow, D. M., Birktoft, J. J., & Hartley, B. S. (1969) *Nature 221*, 337–340.
Bock, K. (1983) *Pure Appl. Chem. 55*, 605–622.
Bode, W., Papamokos, E., & Musil, D. (1987) *Eur. J. Biochem. 166*, 673–692.
Breddam, K. (1986) *Carlsberg Res. Commun. 51*, 83–128.
Breddam, K. (1988) *Carlsberg Res. Commun. 53*, 309–328.
Breddam, K., Sorensen, S. B., & Svendsen, I. (1987) *Carlsberg Res. Commun. 52*, 297–311.
Brenner, S. (1988) *Nature 334*, 528–530.
Carter, P., & Wells, J. A. (1988) *Nature 332*, 564–568.
Carter, P., & Wells, J. A. (1990) *Proteins 7*, 335–342.
Cohen, G. H., Silvertion, E. W., & Davies, D. R. (1981) *J. Mol. Biol. 148*, 449–479.
Cooper, A., & Bussey, H. (1989) *Mol. Cell. Biol. 9*, 2706–2714.
Derewenda, Z. S., & Derewenda, U. (1991) *Biochem. Cell Biol. 69*, 842–851.
Dixon, M. M., & Matthews, B. W. (1989) *Biochemistry 28*, 7033–7038.
Douglass, J., Civelli, O., & Herbert, E. (1984) *Annu. Rev. Biochem. 53*, 665–715.
Franken, S. M., Rozeboom, H. J., Kalk, K. H., & Dijkstra, B. W. (1991) *EMBO J. 10*, 1297–1302.
Galjart, N. J., Gillemans, N., Harris, A., van der Horst, G. T. J., Verheijen, F. W., Galjaard, H., & d'Azzo, A. (1988) *Cell 54*, 755–764.
Hayashi, R., Moore, S., & Stein, W. H. (1973) *J. Biol. Chem. 248*, 8366–8369.

Hirono, S., Akagawa, H., Mitsui, Y., & Iitaka, Y. (1984) *J. Mol. Biol. 178*, 389–413,
Howard, A. J., Nielsen, C., & Xuong, N. H. (1985) *Methods Enzymol. 114*, 452–471.
Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H., & Salemme, F. R. (1987) *J. Appl. Crystallogr. 20*, 383–387.
Jones, T. A., & Thirup, S. (1986) *EMBO J. 5*, 819–822.
Jones, W. M., Scaloni, A., Bossa, F., Popowicz, A. M., Schneewind, O., & Manning, J. M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2194–2198.
Kabsch. W. (1978) *Acta Crystallogr. A34*, 827–828.
Kabsch, W. (1988) *J. Appl. Crystallogr. 21*, 916–924.
Kossiakof, A. A., & Spencer, S. A. (1981) *Biochemistry 20*, 6462–6474.
Kraut, J., Robertus, J. D., Birktoft, J. J., Alden, R. A., Wilcox, P. E., & Powers, J. C. (1972) *Cold Spring Harbor Symp. Quant. Biol. 36*, 117–123.
Lennarz, W. J. (1980) *The Biochemistry of Glycoproteins and Proteoglycans*, pp 20–21, Plenum Press, New York.
Leslie, A. W. (1987) *Acta Crystallogr. A43*, 134–136.
Lewis, P. N., Momany, F. A., & Scheraga, H. A. (1973) *Biochim. Biophys. Acta 303*, 211–229.
Liao, D.-I., & Remington, S. J. (1990) *J. Biol. Chem. 265*, 6528–6531.
McGrath, M. E., Vasquez, J. R., Craik, C. S., Yan, A. S., Honig, B., & Fletterick, R. J. (1992) *Biochemistry 31*, 3059–3064.
McPhalen, C. A., & James, M. N. G. (1988) *Biochemistry 27*, 6582–6598.
Masserschmidt, A., & Pflugrath, J. W. (1987) *J. Appl. Crystallogr. 20*, 306–315.
Miller, S., Janin, J., Lesk, A. M., & Chothia, C. (1987) *J. Mol. Biol. 196*, 641–656.
Odya, C. E., & Erdös, E. G. (1981) *Methods Enzymol. 80*, 460–466.
Odya, C. E., Marinkovic, D. B., Hammon, K. J., Stewart, T. A., & Ordos, E. G. (1978) *J. Biol. Chem. 253*, 5927–5931.
Ollis, D. L., Cheah, E., Cygler, M., Dykstra, B., Frolow, F., Fraken, S., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J., & Goldman, A. (1992) *Protein Eng.* (in press).
Pathak, D., Ngai, K. L., & Ollis, D. (1988) *J. Mol. Biol. 204*, 435–445.
Polgar, L. (1991) *Eur. J. Biochem. 197*, 441–447.
Ramachandran, G., & Sasisekharan, U. (1968) *Adv. Protein Chem. 23*, 283–437.
Rawlings, N. D., Polgar, L., & Barrett, A. J. (1991) *Biochem. J. 279*, 907–911.
Remington, S. J., Wiegand, G., & Huber, R. (1982) *J. Mol. Biol. 158*, 111–152.
Schecter, I., & Berger, B. (1967) *Biochem. Biophys. Res. Commun. 27*, 157–162.
Schmidt, M. F., Weaver, L. H., Holmes, M. A., Gruetter, M. G., Ohlendorf, D. H., Reynolds, R. A., Remington, S. J., & Matthews, B. W. (1981) *Acta Crystallogr. A37*, 701–710.
Schrag, J. D., Li, Y., Wu, S., & Cygler, M. (1991) *Nature 351*, 761–764.
Segal, D. M., Cohen, G. H., Davies, D. R., Powers, J. C., & Wilcox, P. E. (1971) *Cold Spring Harbor Symp. Quant. Biol. 36*, 85–90.
Sorenson, S. B., Svendsen, I., & Breddam, K. (1986) *Carlsberg Res. Commun. 52*, 297–311.
Sprang, S., Standing, T., Fletterick, R. J., Stroud, R. M., Finer-Moore, J., Xuong, N. H., Hamlin, R., Rutter, W. J., & Craik, C. S. (1987) *Science 237*, 905–909.
Steigemann, W. (1974) Doctoral Thesis, Technical University, Munich.
Steiner, D. F., Quinn, P. S., Chan, S. J., Marhs, J., & Tager, H. S. (1980) *Ann. N.Y. Acad. Sci. 343*, 1–6.
Sussman, J. L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L., & Silman, I. (1991) *Science 253*, 872–879.
Thomas, L., Cooper, A., Bussey, H., & Thomas, G. (1990) *J. Biol. Chem. 265*, 10821–10824.
Tronrud, D. E., TenEyck, L. F., & Matthews, B. W. (1987) *Acta Crystallogr. A43*, 489–501.
Wang, B. C. (1985) *Methods Enzymol. 115*, 90–112.
Warshel, A., & Russel, S. T. (1986) *J. Am. Chem. Soc. 108*, 6569–6579.
Warshel, A., Naray-Szabo, G., Sussman, F., & Hwang, J.-K. (1989) *Biochemistry 28*, 3629–3637.
Wilson, I. A., Skehel, J. J., & Wiley, D. C. (1981) *Nature 289*, 266–273.
Wilson, K. P., Liao, D.-I., Bullock, T., Remington, S. J., & Breddam, K. (1990) *J. Mol. Biol. 221*, 301–303.
Wright, C. S. (1972) *J. Mol. Biol. 67*, 151–163.
Zhou, X. Y., Galjart, N. J., Willemsen, R., Gillemans, N., Galjaard, H., & d'Azzo, A. (1991) *EMBO J. 10*, 4041–4048.
Zuber, H., & Matile, P. H. (1968) *Z. Naturforsch, 23B*, 663–665.