2.8-Å Structure of Yeast Serine Carboxypeptidase[†,‡]

James A. Endrizzi,[§] Klaus Breddam,[∥] and S. James Remington*[,⊥]

Institute of Molecular Biology and Departments of Chemistry and Physics, University of Oregon, Eugene, Oregon 97403, and Department of Chemistry, Carlsberg Laboratory, Gamle Carlsbergvej 10, DK-2500 Copenhagen Valby, Denmark

Received February 23, 1994; Revised Manuscript Received June 10, 1994[⊗]

ABSTRACT: The structure of monomeric serine carboxypeptidase from *Saccharomyces cerevisiae* (CPD-Y), deglycosylated by an efficient new procedure, has been determined by multiple isomorphous replacement and crystallographic refinement. The model contains 3333 non-hydrogen atoms, all 421 amino acids, 3 of 4 carbohydrate residues, 5 disulfide bridges, and 38 water molecules. The standard crystallographic *R*-factor is 0.162 for 10 909 reflections observed between 20.0- and 2.8-Å resolution. The model has rms deviations from ideality of 0.016 Å for bond lengths and 2.7° for bond angles and from restrained thermal parameters of 7.9 Å^2. CPD-Y, which exhibits a preference for hydrophobic peptides, is distantly related to dimeric wheat serine carboxypeptidase II (CPD-WII), which has a preference for basic peptides. Comparison of the two structures suggests that substitution of hydrophobic residues in CPD-Y for negatively charged residues in CPD-WII in the binding site is largely responsible for this difference. Catalytic residues are in essentially identical configurations in the two molecules, including strained main-chain conformational angles for three active site residues (Ser 146, Gly 52, and Gly 53) and an unusual hydrogen bond between the carboxyl groups of Glu 145 and Glu 65. The binding of an inhibitor, benzylsuccinic acid, suggests that the C-terminal carboxylate binding site for peptide substrates is Asn 51, Gly 52, Glu 145, and His 397 and that the "oxyanion hole" consists of the amides of Gly 53 and Tyr 147. A surprising result of the study is that the domains consisting of residues 180–317, which form a largely α-helical insertion into the highly conserved cores surrounding the active site, are quite different structurally in the two molecules. It is suggested that these domains have evolved much more rapidly than other parts of the molecule and are involved in substrate recognition.

Serine carboxypeptidases are exopeptidases that efficiently remove C-terminal amino acids from peptides. They hydrolyze peptide esters and will also release ammonia from N-blocked C-termini [for reviews see Breddam (1986) and Remington and Breddam (1994)]. The enzymes are inhibited by diisopropyl phosphofluoridate, a reagent specific for serine proteinases, suggesting that they have a Ser-His-Asp catalytic triad as found in the trypsin and chymotrypsin families of endopeptidases. Chemical modifications of the vacuolar enzyme from yeast (CPD-Y)[1] identified an essential serine (Ser 146) (Hayashi et al., 1973) and implicated a histidine (Hayashi et al., 1975), later shown by mutagenesis to be His 397 (Bech & Breddam, 1989). Structural studies of a distantly related enzyme, serine carboxypeptidase II from wheat (CPD-WII) identified Asp 338 as a third member of the triad (Liao & Remington, 1990). Thus, it seems likely that the enzymatic mechanism is similar to that of the serine endopeptidases. However, serine carboxypeptidases have peptidase activity which is optimal at pH 4.5–5.5, depending on the enzyme, in contrast to the serine endopeptidases, which are inactive below pH ~7.

Serine carboxypeptidases are found in every eukaryote examined and fall into three general classes based on substrate specificity. Highly specific serine carboxypeptidases have been shown to be components of the peptide hormone maturation machinery (Cooper & Bussey, 1989) and also degrade peptide hormones; for example, kidney prolyl carboxypeptidase hydrolyzes angiotensins II and III and may be involved in regulation of blood pressure (Odya & Erdös, 1981). The amino acid sequence of the latter enzyme has recently been determined, revealing homology with both serine carboxypeptidases and prolyl endopeptidase (Tan et al., 1993a) and providing evidence that a serine endopeptidase and exopeptidase can have a common structural framework.

The less specific serine carboxypeptidases are grouped into two general classes with some overlap, recently termed carboxypeptidases C and D. Carboxypeptidases C are subsets of the serine carboxypeptidase family with high specificity for hydrophobic residues at the P_1' position [the C-terminal residue; notation of Schechter and Berger (1967)] while carboxypeptidases D most efficiently hydrolyze basic residues at P_1'. The three-dimensional structure of one member of the latter family, CPD-WII, has been refined at 2.2-Å resolution (Liao et al., 1992). CPD-Y is a member of the former family and is the subject of this investigation. In this report, we describe an atomic model of CPD-Y determined by multiple isomorphous replacement and refined at 2.8-Å resolution. The structures of its complexes with benzylsuccinic acid and *p*-(chloromercuri)benzoate have also been determined. The models are compared with that of CPD-WII in order to gain insight into the molecular basis for the difference in substrate specificities.

MATERIALS AND METHODS

Deglycosylation/Purification. CPD-Y was efficiently deglycosylated using endoglycosidase H (endo-H, Boehringer)

[†] This work was supported in part by grants from the National Science Foundation (MCB 9118302 to S.J.R.) and to the Institute of Molecular Biology from the Lucille P. Markey Foundation.

[‡] The model of CPD-Y has been deposited in the Brookhaven Protein Data Bank under the filename 1YSC.

* Address correspondence to this author at the Institute of Molecular Biology [telephone (503) 346-5151].

[§] Institute of Molecular Biology and Department of Chemistry, University of Oregon.

[∥] Carlsberg Laboratory.

[⊥] Institute of Molecular Biology and Department of Physics, University of Oregon.

[⊗] Abstract published in *Advance ACS Abstracts,* August 15, 1994.

[1] Abbreviations: CPD-Y, yeast serine carboxypeptidase; CPD-WII, serine carboxypeptidase II from wheat bran; BZS, benzylsuccinic acid; MIR, multiple isomorphous replacement; rms, root mean square; pCMB, *p*-(chloromercuri)benzoate; MMI, methylmercury iodide.

0006-2960/94/0433-11106$04.50/0

by a variation on the manufacturer's recommended procedure. It was determined that the reaction is roughly 100-fold more efficient at pH 4.5 in acetate buffer as opposed to pH 5.5 in citrate buffer. Preparative scale deglycosylation reactions, which would have required prohibitive amounts of endo-H using the recommended conditions, consisted of 11 mL of CPD-Y (18.8 mg/mL in water), 5 mL of 0.05 M sodium acetate (pH 4.5), 27 mL of water, and 0.1 unit of endo-H in 0.1 mL of water. The reaction was allowed to proceed for 24 h at 35 °C. During the reaction, 5-μL aliquots were diluted with 250 μL of water for activity assays. CPD-Y activity was assayed spectroscopically at 340 nm with the following reaction mixture: 965 μL of 0.05 M MES and 1 mM EDTA, pH 6.5, + 25 μL of 8 mM furylacryloyl-Phe-Leu-OH in methanol + 10 μL of CPD-Y.

The deglycosylated enzyme was repurified by affinity chromatography as previously described (Johansen et al., 1976). CPD-Y fractions were pooled and dialyzed against water. Sodium citrate (10 mg), pH 5.3, was added per milligram of CPD-Y, followed by lyophilization. The molecular weight was estimated by SDS–polyacrylamide gel electrophoresis using the Pharmacia PHAST system after reduction with dithiothreitol.

Crystallization. Crystals were grown by hanging drop vapor diffusion in tissue culture plates (Linbro). A total of 5 μL of well solution consisting of 18%–24% poly(ethylene glycol) (PEG), M_r = 6000, 0.3 M NaOAc, 0.05 M NaCl, and 100 mM imidazole/NaOH, pH 6–8, was added to 5 μL of CPD-Y at 10 mg/mL in 0.1 M NaCl, 1 mM DTT, and 20 mM citrate/NaOH, pH 6.5, on a silanized cover slip and inverted over wells containing 1 mL of precipitant. At room temperature orthorhombic crystals grow within days but are fragile and sensitive to X-rays. After several months, the same drops produce a cubic crystal form that is more stable in the X-ray beam. Data quality cubic crystals can be grown in 2 weeks by microseeding drops with fragments of crushed crystals. Crystals were stored in a solution identical to that in the well, but containing 26%–28% PEG-6000.

Data Collection. Data were collected at room temperature on a San Diego Multiwire Systems area detector using graphite-monochromated Cu Kα from a Rigaku RU200 rotating anode operated at 40 kV and 150 mA. Counts were summed over 0.08–0.12° ranges in Ω for exposure times of 60 s or longer. The data were reduced assuming 222 symmetry with the supplied software (Howard et al., 1985), which in its present form cannot handle cubic space groups, followed by reduction to a $P2_13$ asymmetric unit with the PROTEIN program (Steigemann, 1974). The space group and unit cell dimensions were initially determined by analysis of precession photographs.

Heavy Atom Derivatives. Potential heavy atom derivatives were screened by soaking crystals in storage solution (without DTT) containing 1%, 10%, and 100% saturated heavy atom compound for 1 day to 2 weeks, followed by precession photography. Complete data sets were taken on heavy atom derivatives pCMB, methylmercury iodide (MMI), and $PtCl_4$. The following conditions were used: pCMB (1 mM, 3 days), $PtCl_4$ (10 mM, overnight), and MMI (saturated, 4 days).

Phase Determination and Model Building. Difference Patterson maps were readily interpretable in terms of a single site for each of the three derivatives, and parameters were refined with the phase refinement module of PROTEIN. Difference Fourier methods were used to look for minor binding sites. Heavy atom parameters were refined separately to convergence and then were combined. An electron density map was calculated at 3.5-Å resolution, which immediately revealed helical segments of the correct handedness. The α-carbon coordinates of CPD-WII (Liao et al., 1992) were positioned approximately into the map by hand using FRODO (Jones & Thurip, 1986) and used as a starting point for model building with FRODO. All amino acids and side chains which had appropriate electron density [CPD-Y amino acid sequence from Valls et al. (1987)] were included in the initial model, which contained 293 out of 421 amino acids but did not include most of residues 180–317. At various stages, model phases were combined with heavy atom phases using the procedure described by Hendrickson and Lattman (1970), and $2F_o - F_c$ maps were inspected. In later stages $2F_o - F_c$ maps using model phases were inspected, as were $F_o - F_c$ maps.

Crystallographic Refinement. The atomic model was refined with the TNT package (Tronrud et al., 1987) using "conjugate direction" function minimization (Tronrud, 1992). The model was subjected to 10–30 cycles of TNT refinement after each round of model building. Initially, coordinate refinement was performed with a single overall temperature factor of 20.0. Correlated *B*-value refinement was performed after the model was nearly complete, and phases were extended to 2.8 Å.

RESULTS

Deglycosylation. CPD-Y contains four N-linked carbohydrate chains (Ballou et al., 1990). Using the optimized deglycosylation conditions discussed, SDS gel electrophoresis indicated that fully glycosylated CPD-Y (M_r 66 000) was converted to products of molecular weight 57 000 and 53 000 within 1 h, indicating stepwise removal of carbohydrate. After 24 h, only the 53 000 molecular weight band remained, corresponding to a reduction in molecular mass of 13 000 Da. The activity of CPD-Y remained constant through the course of deglycosylation, suggesting that the active site conformation is not affected by removal of carbohydrate. The final yield of deglycosylated product was 70%.

Crystallization. Deglycosylated CPD-Y has been crystallized in two forms. Orthorhombic crystals of space group $C222$, cell dimensions a = 161.5 Å, b = 137.9 Å, and c = 87.5 Å, grow within a week but are sensitive to X-rays at room temperature. They have a packing parameter, V_m, of 2.4 Å^3/Da calculated for a dimer in the asymmetric unit (Matthews, 1968). After several months, cubic $P2_13$ crystals with cell dimension a = 111.8 Å grow out of the same drops. Seeding results in larger single crystals. $P2_13$ crystals have a monomer in the asymmetric unit, with V_m = 2.1Å^3/Da. Both crystal forms diffract to about 2.5 Å, but the cubic crystals were chosen for further study since they are less sensitive to X-rays at room temperature and have higher symmetry and smaller cell dimensions relative to the orthorhombic crystals.

Data Collection. Diffraction data sets were collected on a San Diego Multiwire Systems area detector from single crystals of native CPD-Y, three heavy atom derivatives, and CPD-Y complexed with benzylsuccinic acid. The signal to noise ratio was 2:1 or greater on intensities for native and pCMB data to 2.8 Å, CPD-Y/benzylsuccinic acid data to 3.2 Å, and K_2-$PtCl_4$ and MMI data to 3.5 Å. Data collection statistics are given in Table 1.

Phase Calculation. Heavy atom parameters and their reaction sites are given in Table 2. Earlier work had shown that pCMB covalently attaches to Cys 341 with 1:1 stoichiometry and alters substrate specificity at the S_1 subsite (Breddam & Svendsen, 1984), so it was an obvious choice. Patterson maps were readily interpretable in terms of a single site for each derivative, and difference Fourier maps did not

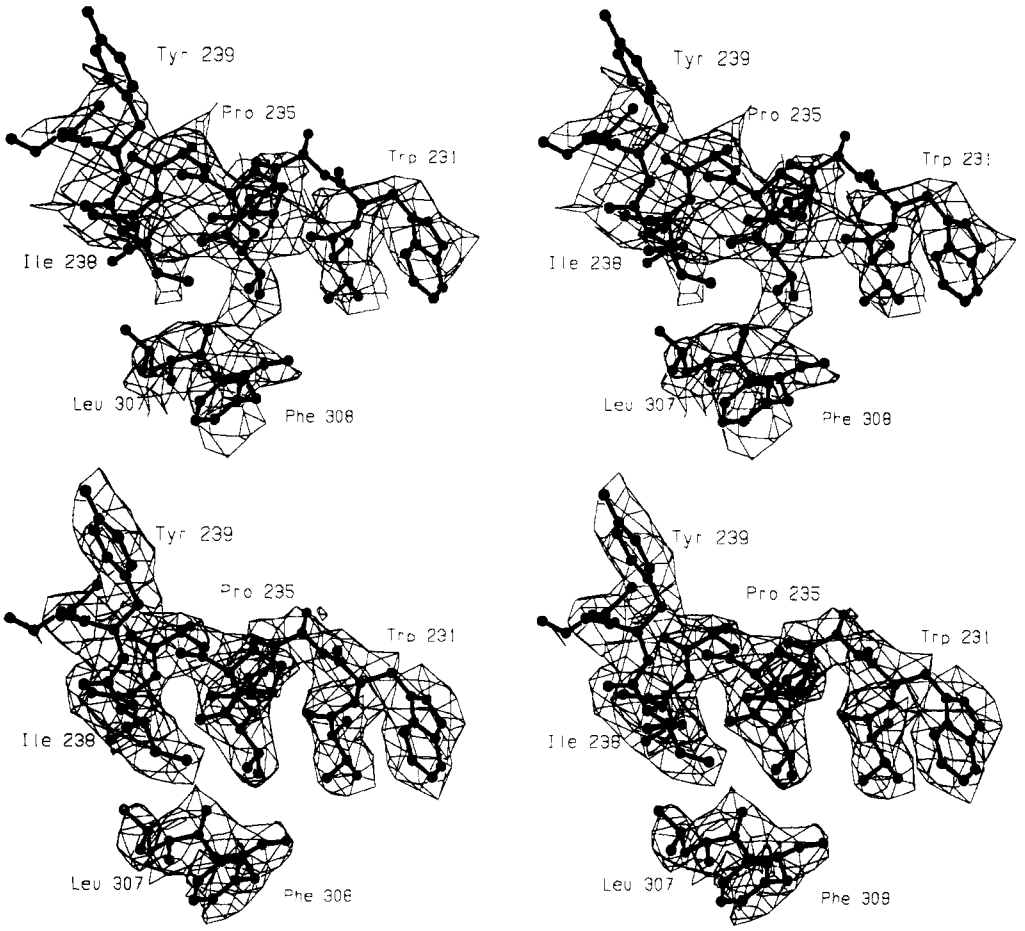

FIGURE 1: 3.5-Å MIR electron density (top) and the final 2.8-Å $2F_o - F_c$ map (bottom), both contoured at 1 SD. Part of the final model of CPD-Y in a region where there are two insertions relative to CPD-WII is superimposed.

reveal minor sites. Anisotropic temperature factors were refined after positional refinement had converged. Heavy atom parameters are given in Table 2. pCMB is an excellent derivative, $PtCl_4$ is of intermediate quality, and MMI is weak, partly due to its location on a crystallographic 3-fold. The initial 3.5-Å electron density map had an overall figure of merit = 0.50.

Model Building/Refinement. Inspection of the original 3.5-Å MIR phased difference map revealed several helices which were used as a guide to manually position the model of the enzyme from wheat (CPD-WII). Electron density was verified for the following features: single carbohydrate residues at Asn 87, Asn 168, and Asn 368; the catalytic triad and flanking residues; five disulfide bonds; and several model insertions in CPD-Y relative to CPD-WII. The starting model, containing 2241/3338 protein atoms, had an initial R-factor of 0.438 for all data from 8 to 3.5 Å. After refinement, model phases were combined with the heavy atom derivatives, resulting in improved electron density maps. Several rounds of model building into MIR and phase combined maps followed by positional refinement were performed until the model was approximately 85% complete on an atom basis, after which point the heavy atom derivative data were not used. Omit maps were inspected for ambiguous segments of the peptide chain. With 96% of the protein atoms included in the model, atoms in peaks above 3σ in the $F_o - F_c$ map that could make at least one hydrogen bond to protein were included as water molecules. Figure 1 illustrates the improvement in the quality of the electron density between the 3.5-Å MIR map and the final 2.8-Å $2F_o - F_c$ map for several turns of helix not present in the structure of CPD-WII.

Table 1: Data Collection Statistics

data set	crystals	resolution (Å)	reflections total (unique)	R_{merge}/R_{iso}[a] (%)	% complete
native	1	2.8	46 348 (10 909)	4.3/–	93
pCMB	1	2.8	64 641 (10 487)	7.5/17.1	89
$PtCl_4$	1	3.5	34 943 (5 803)	6.5/15.9	96
MMI	1	3.5	31 776 (5 227)	8.6/17.9	86
BZS	1	3.2	69 291 (7 748)	9.9/16.7	98

[a] $R_{merge} = \sum_{hkl}\sum_i|I^i_{hkl} - \langle I_{hkl}\rangle|/\sum_{hkl}\sum_i I^i_{hkl}$. $R_{iso} = \sum_{hkl}|I_{der} - I_{nat}|/\sum_{hkl} I_{nat}$.

Table 2: Heavy Atom Parameters

compd	sites	X	Y	Z	occupancy[a]	ligand
pCMB	1	0.454	0.143	0.594	1.60	Cys 341
$PtCl_4$	1	0.674	0.004	0.529	1.60	disulfide 56–298
MMI	1	0.025	0.025	0.025	0.79	Phe 308 (3-fold)

[a] Occupancy is on an arbitrary scale.

The final model consists of 3333 non-hydrogen atoms, all 421 amino acids, 3 of 4 N-linked carbohydrate residues, and 38 water molecules and has an R-factor of 0.162 for 10 909 independent reflections between 20.0 and 2.8 Å (a solvent model was included, with $K_{sol} = 0.6$ and $B_{sol} = 514.0$). The final model has rms deviations from ideality of 0.016 Å for bond lengths, 2.7° for bond angles, and 0.01 Å for planar groups. The rms B-value difference for covalently linked atoms is 7.9 Å^2.

Molecular Architecture. As expected, the fold of CPD-Y is very similar to that previously described for CPD-WII (Laio et al., 1992) and other members of the "α/β hydrolase-fold" family of enzymes (Ollis et al., 1992). The molecule has a

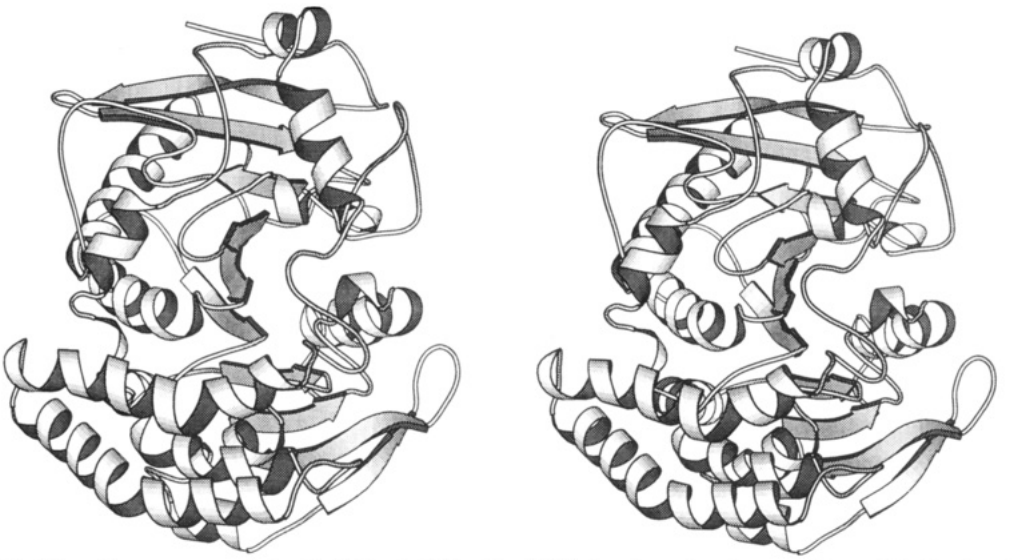

FIGURE 2: CPD-Y ribbon diagram generated with Molscript (Kraulis, 1991) based on α-carbon positions of the final model.

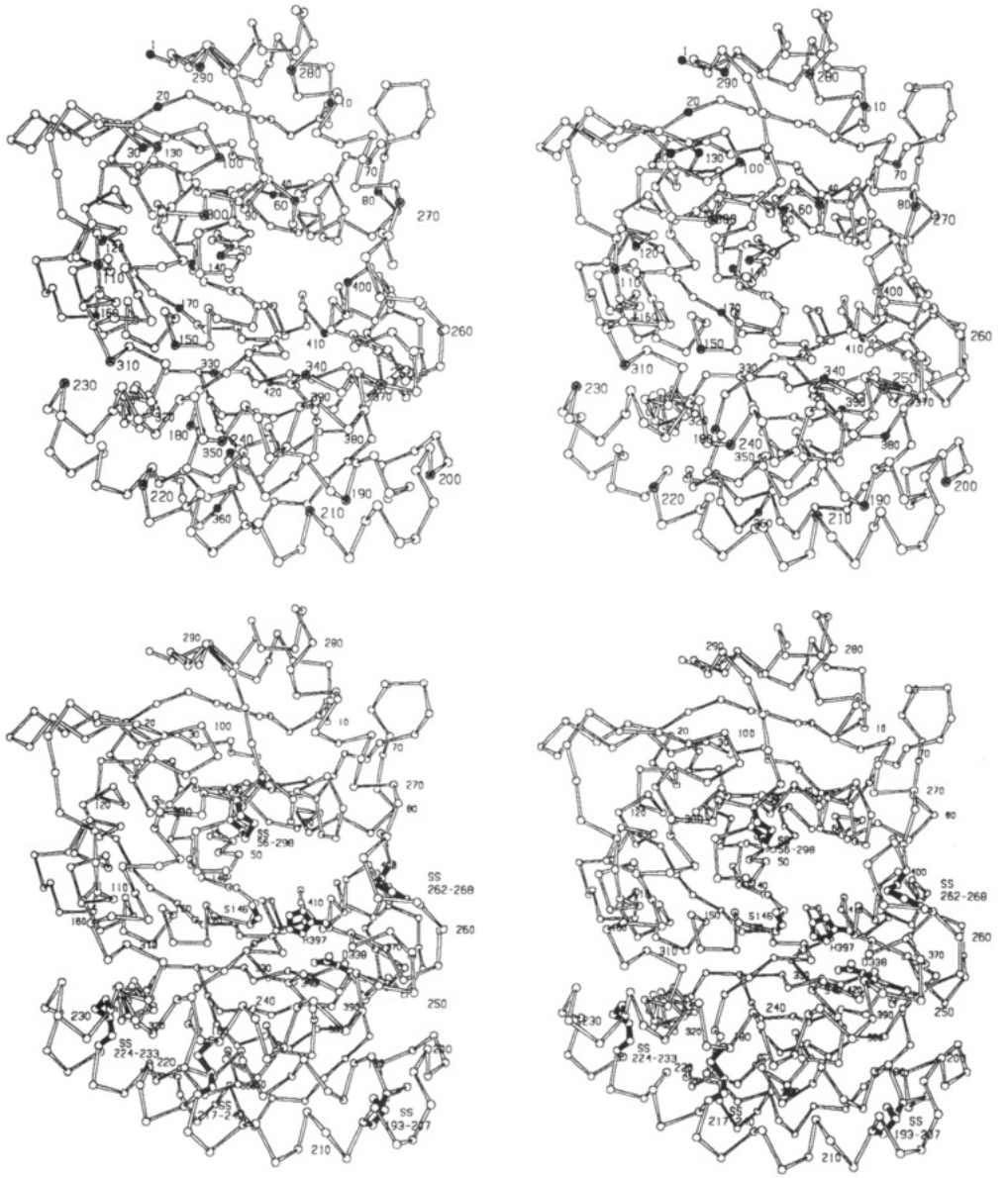

FIGURE 3: (a, top) Stereo drawing of the α-carbon backbone of CPD-Y. Every tenth residue is labeled and the corresponding atom is filled. (b, bottom) Stereo drawing of the α-carbon backbone of CPD-Y, including atoms of the catalytic triad and all five disulfide bonds.

central 11-stranded mixed β-sheet that twists approximately 180° end to end (the N-terminal strand is short and irregular) and is flanked by 14 α-helices (Figures 2 and 3). The central 6 strands of β-sheet (5–9) are parallel with the serine 146

nucleophile located at an "elbow" between β-strand 6 and the following α-helix (Figure 3b). Disulfide linkages are found between residues 56–298, 193–207, 217–240, 224–233, and 262–268 and are located on one face of the molecule surrounding the active site cavity (Figure 3b). Disulfide 262–268, which is in a high-energy conformation, has electron density suggesting that this disulfide is partially reduced or "mobile". Pro 54 and Pro 96 are in the cis configuration. Of the four glycosylation sites, Asn 87, Asn 168, and Asn 368 each appear to have a single N-linked *N*-acetylglucosamine residue, but Asn 13 has no electron density attributable to carbohydrate, presumably due to mobility. None of the glycosylation sites are near the active site, consistent with the fact that glycosylation is not required for *in vivo* stability or activity of CPD-Y (Winther et al., 1991).

The 15 N-terminal residues and helix 282–288, which pack against each other, appear to be flexible and have high *B*-values, averaging about 60 Å^2. The side chains of residues 1–4, 6–7, 13–17, 24, 26, 35, 43, 69, 86, 90, 204, 213, 232, 281, 284, 291, 359, 364, 365, and 372 have little or no electron density and were modeled as alanine. Secondary structural assignments of CPD-Y based on the method of Kabsch and Sander (1983) are shown in Figure 4 and compared with the secondary structure of CPD-WII.

A Ramachandran plot for main-chain (Φ,Ψ) torsion angles (Figure 5) has few residues outside the "allowed" low-energy conformations. Some of these are in loop regions which have weak electron density, but as seen in the 2.2-Å model of CPD-WII (Liao et al., 1992), active site residues Gly 52, Gly 53, and Ser 146 have well-defined electron density but (Φ,Ψ) values in high-energy regions of the diagram. In fact, the α-carbons of Gly 53 and Pro 54 have a very short contact distance of 2.7 Å in the final model. Residues Gly 100 and Lys 385, located in tight turns, also have high-energy main-chain torsion angles and well defined electron density.

Active Site. Residues comprising the catalytic triad (Ser 146-Asp 338-His 397) lie near the bottom of a large hydrophobic pit and are clearly defined in the electron density map (Figure 6). Near the catalytic residues are two hydrophobic pockets, which could accommodate hydrophobic side chains at the P_1 and P_1' positions (Figure 7). A comparison of active site features of CPD-Y and CPD-WII are presented in the Discussion.

CPD-Y/Benzylsuccinic Acid (BZS) Complex. Crystals of CPD-Y were soaked with 10 mM BZS in standard storage solution at pH 4.5 for 2 weeks before data collection. A model was constructed and refined to an *R*-factor of 16.2% for all data between 20- and 3.2-Å resolution. The model has rms deviations from ideal bond lengths and angles of 0.019 Å and 3.0°, respectively. The final $2F_o - F_c$ map shows electron density for the entire inhibitor (Figure 8a). The α- and β-carboxylates of BZS occupy the carboxyl-terminal binding site and "oxyanion hole" of CPD-Y, respectively (Figure 8b). Atom O2A of the α-carboxylate forms hydrogen bonds with Asn 51 ND2 (2.7 Å), the backbone amide of Gly 52 (2.8 Å), and Glu 145 OE1 (3.3 Å), while atom O2B has hydrogen bonds to Glu 145 OE1 (2.6 Å) and His 397 NE2 (3.4 Å). Atom OD1 of the β-carboxylate has hydrogen bonds to the amide of Gly 53 (2.8 Å) and Ser 146 OG (2.9 Å) and is near the amide of Tyr 147 (3.9 Å), while atom OD2 has hydrogen bonds with His 397 NE2 (2.8 Å) and Ser 146 OG (2.90 Å). Atom OG of Ser 146 is also 2.8 Å from His 397 NE2 and 3.1 Å from N147, resulting in four atoms in hydrogen-bonding distance to Ser 146 in the complex. In the final model, the χ_1 rotamer of the catalytic serine rotates from 95° in unliganded CPD-Y to −162° in the complex (Figure 8b).

The phenyl ring of BZS occupies a hydrophobic cavity, presumably the S_1' subsite, with Thr 60, Phe 64, Glu 65, Tyr 256, Tyr 269, Leu 272, Met 398, and disulfide 56–298 all within 5.5 Å (Figure 8b). Methionine 398 SD moves over 1 Å relative to its position in wild-type CPD-Y in response to BZS binding, as it would sterically overlap with the phenyl ring in the unliganded conformation. Also in response to BZS binding, Thr 60 appears to adopt a new χ_1 rotamer (9°), (−42° in unliganded CPD-Y).

pCMB-Modified CPD-Y. A model of pCMB-modified CPD-Y was refined to an *R*-factor of 0.17 for all data from 20.0 to 2.8 Å, with deviations from ideality of 0.017 Å and 2.8° for bonds and angles, respectively. Since it is known that reaction of pCMB modifies the specificity of CPD-Y at P_1 toward smaller side chains, the model suggests the approximate location of the S_1 subsite. The phenyl ring of pCMB lies in a hydrophobic depression formed by Tyr 147, Leu 178, Leu 245, Trp 312, Ile 340, and Cys 341 (Figure 9). In addition, the carboxylate of pCMB is hydrogen bonded to the side chain of Trp 231 of a symmetry-related molecule (Figure 9).

DISCUSSION

Comparison of CPD-Y and CPD-WII Overall Fold. CPD-Y and CPD-WII are distantly related enzymes, with only 26% amino acid sequence identity. In solution, CPD-Y is a single-chain monomer, while CPD-WII is a homodimer, each subunit consisting of two polypeptide chains linked by disulfide bonds. The topologies of the two folds are identical. However, some surprising and unexpected structural rearrangements occur that are very likely to be related to differences in substrate specificity and state of oligomerization.

A modification of the procedure of Rossman and Argos (1976) by W. Bennett (OVRLAP) was used to align CPD-Y and CPD-WII amino acid sequences based on α-carbon coordinates. This method aligns sequences by allowing redefinition of equivalences to maximize local structural agreement, thus indicating the locations of insertions and deletions. This method identifies 17 insertions and deletions, some of which give rise to secondary structural transitions (Figure 4 and below). The rms deviation for 342 topologically equivalent α-carbon positions (out of approximately 420) using this procedure is 1.73 Å [E1, E2 = 2.0 Å; see Rossmann and Argos (1976) for definition]. As noted by Lesk and Chothia (1980) for comparison of related hemoglobins, repacking of conserved secondary structural elements can cause them to translate by as much as 5 Å. Thus, for comparison of CPD-Y and CPD-WII binding sites that follow, residues whose α-carbon coordinates differed by more than 1.5 Å after the superposition described above were excluded from a subsequent superposition. Using these criteria, the rms deviation drops to 0.86 Å for superposition of a highly conserved core of 255 α-carbons (Figure 4), and the sequence identity rises to 90/255 (35%). Secondary structure assignments were obtained by the method of Kabsch and Sander (1983) and are summarized in Figure 4.

Insertion Domain and Hydrophobic Surface. Residues 180–317 of CPD-Y (180–312 of CPD-WII) correspond to an insertion into the basic α/β folding topology of the hydrolase fold family (Figures 2 and 3). This region is largely α-helical, surrounds the active site cavity, and is cross-linked by disulfide bridges (five in CPD-Y and three in CPD-WII). In contrast to the core of the molecule, major structural rearrangements occur in this domain (Figure 10), some of which may be correlated with substrate recognition. Helices are translated

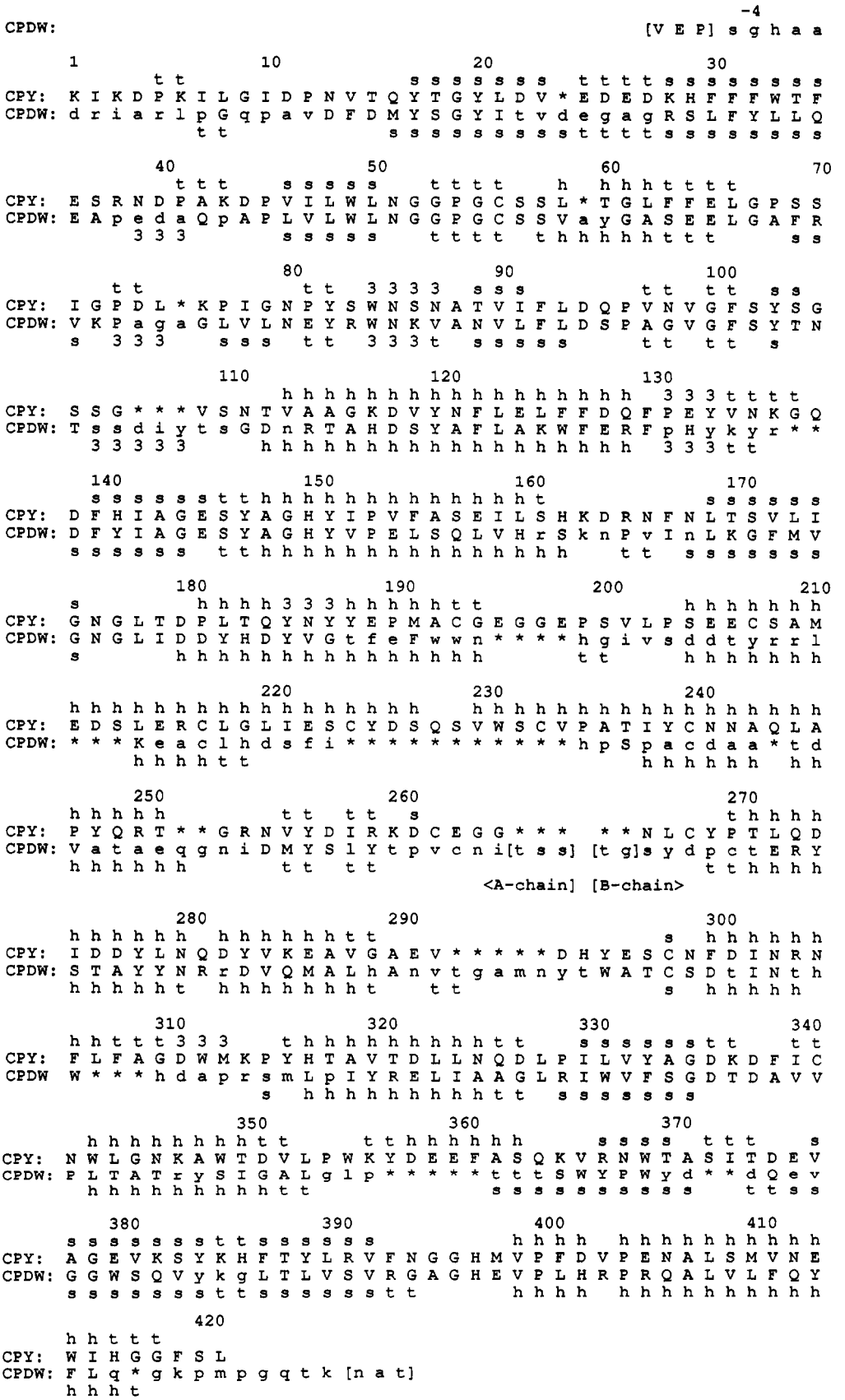

FIGURE 4: Structural alignment of CPD-Y and CPD-WII based on the procedure of Rossmann and Argos (1976). Residues that differ in position by more than 1.5 Å between the two structures are regarded as not structurally conserved, even though the amino acid type might be, and this indicated by lower case lettering in the CPD-WII sequence. Insertion of one amino acid is designated by an asterisk. Secondary structure assignments were determined by the method of Kabsch and Sander (1983) and are designated h (α-helix), 3 (3_{10} helix), s (β-sheet), and t (turn). No electron density is apparent for residues in square brackets. They are not included in the model.

by more than their diameters and change considerably in length. There is no recognizable sequence similarity between residues 180–317 of CPD-Y and 180–312 of CPD-WII, although certain key residues are conserved.

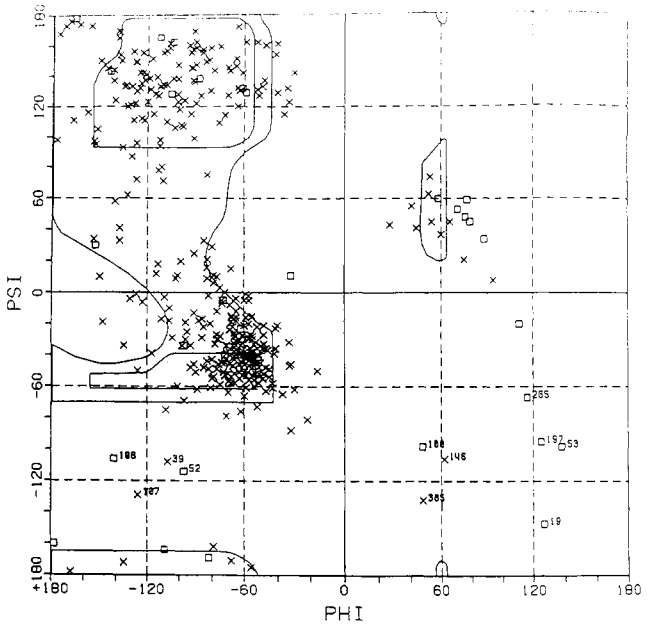

FIGURE 5: Ramachandran plot of main-chain torsional values. Only outliers are labeled. Glycine residues, which are not subject to the same restrictions as other residues, are designated by squares.

Residues 180–234 in CPD-WII comprise a three-helix bundle that forms one wall of the active site cavity. Due to insertion(s) relative to CPD-WII, α-helices 204–227 and 230–251 of CPD-Y are longer than their CPD-WII counterparts by 13 and 8 amino acids, respectively. These elongated α-helices are cross-linked by two disulfide bonds two turns apart, 217–240 and 224–233 (a novel "cysteine zipper"; see Figure 11). In addition, α-helix 300–307 is extended by one turn relative to its CPD-WII homologue. Residues Asp 311, Trp 312, and Met 313 form a new one-turn 3_{10} helix that packs between extended α-helices 230–251 and 300–307.

These new turns of helix form part of an extensive hydrophobic patch near the active site on the protein surface that is not present in CPD-WII. Helix 230–251 contributes Val 230, Trp 231, Val 234, Pro 235, Ile 238, and Tyr 239 to the hydrophobic patch, while helix 300–307 contributes Phe 300, Leu 307, Phe 308, and the aliphatic portion of Arg 304. The "new" 3_{10} helix contributes Trp 312 to this patch. Leucine 221 is also contiguous with this patch, which provides over 750 Å^2 of solvent-exposed hydrophobic surface at the edge of the CPD-Y active site cavity (1.4-Å probe radius (Connolly, 1983); see Figure 7). In fact, this region forms a primary crystal contact. Three molecules interact with these surfaces to form the crystallographic 3-fold axis, primarily by hydrophobic interactions between the extended α-helices and their symmetry mates, but several salt bridges also exist.

We suggest that this hydrophobic patch may be involved in substrate recognition and possibly also "substrate channeling" as suggested by Tan et al. (1993b) for acetylcholinesterase. In the latter enzyme the active site is at the bottom of a funnel lined with hydrophobic and aromatic residues, which they propose functions to increase the on-rate of ligand binding by "electrostatic steering", or directed diffusion. In CPD-WII, this patch is instead acidic and polar, which correlates well with its specificity for basic substrates. Specific hydrophobic (CPD-Y) to acidic or polar (CPD-WII) changes believed to be important for determining substrate specificity are summarized in Table 3. Additional support for the involvement of this patch in substrate recognition comes from alignment of the amino acid sequence of CPD-Y with CPD-S_1 (Svendsen et al., 1993) and CPD-MIII (Sørensen et al., 1989), monomeric serine carboxypeptidases specific for basic and hydrophobic residues, respectively. Model building using CPD-Y as a framework suggests that these enzymes appear to have the three extended helices described above for CPD-Y, including conservation of disulfides 56–298, 217–

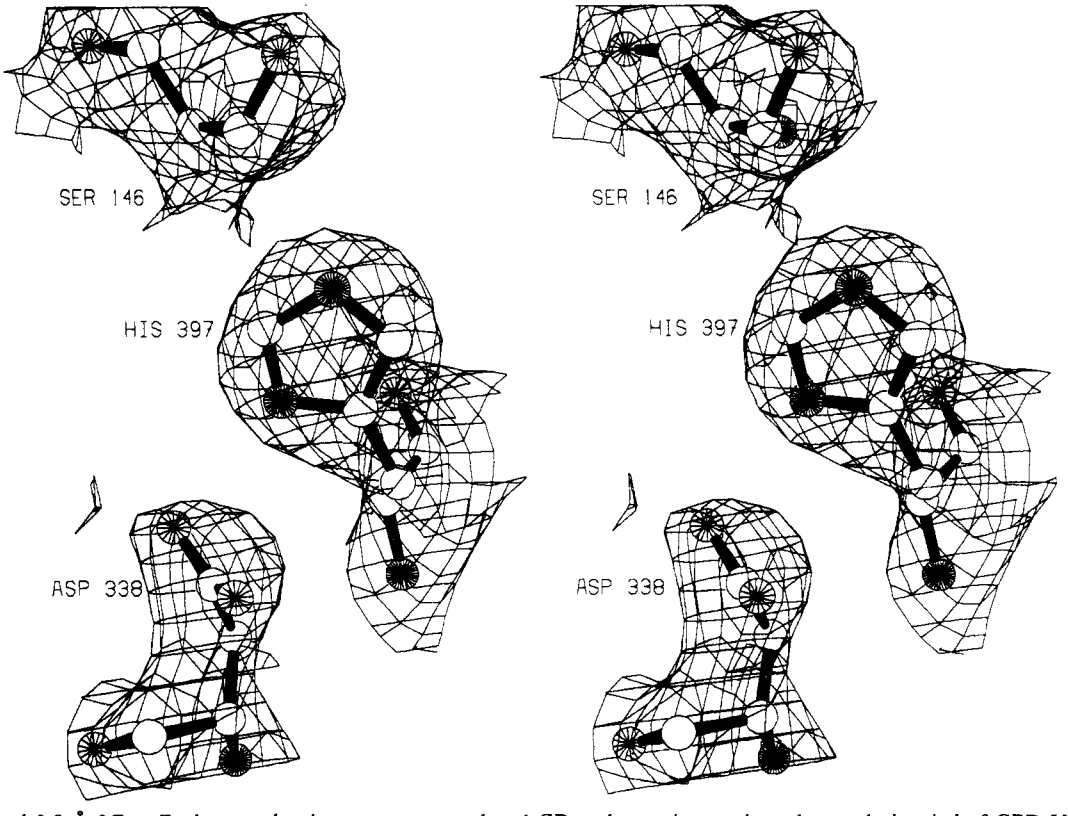

FIGURE 6: Final 2.8-Å $2F_o - F_c$ electron density map contoured at 1 SD and superimposed on the catalytic triad of CPD-Y.

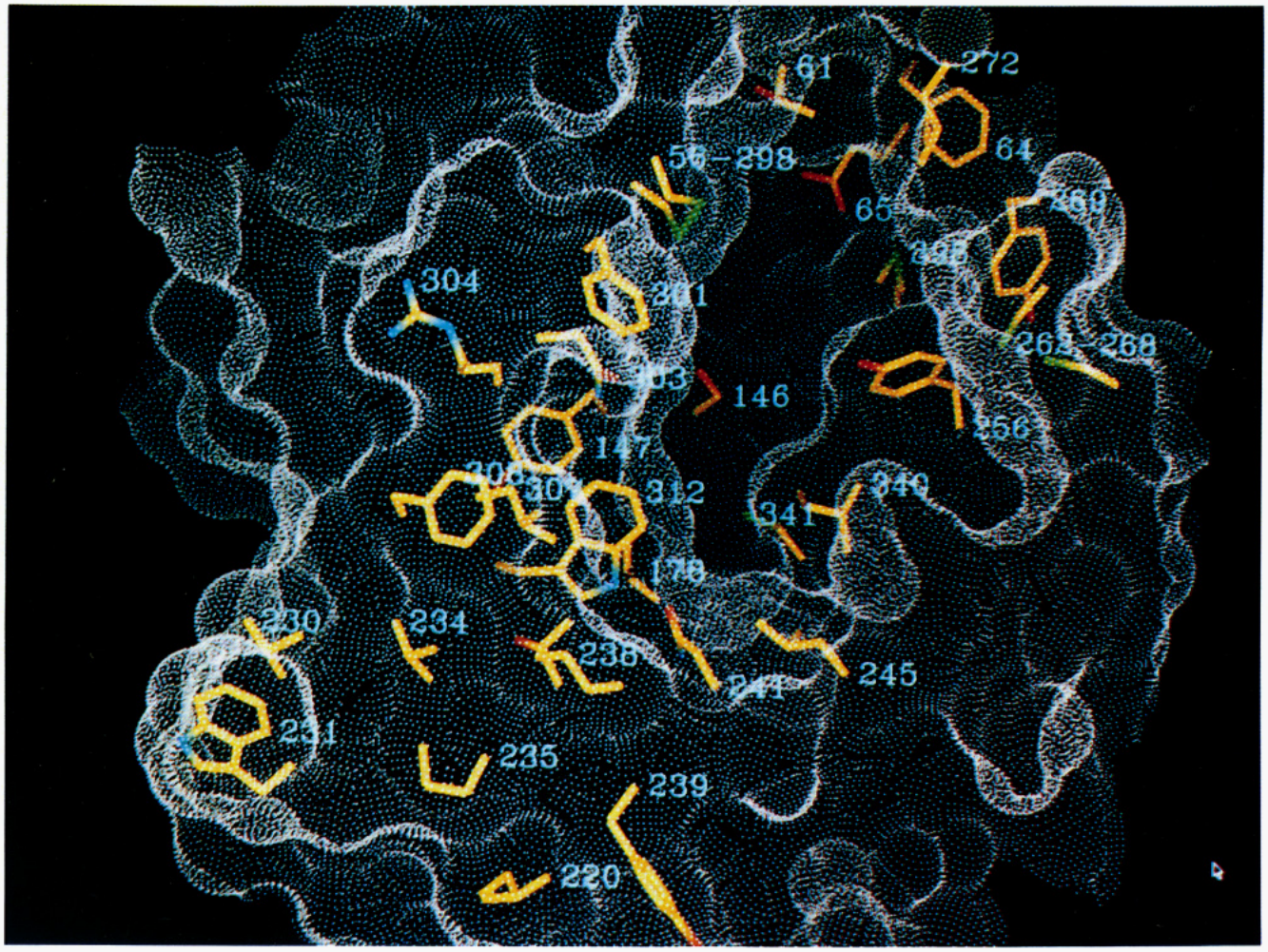

FIGURE 7: Surface area calculated from refined CPD-Y coordinates using the program of Connolly (1983) with a 1.4-Å probe radius. Residues thought to be important for substrate recognition and the serine nucleophile are included and numbered (see text). The extensive solvent-exposed hydrophobic surface at the edge of the active site is located at the lower left of the figure. Note the hydrophobic depressions on either side of Ser 146, representing the S_1' (upper) and and S_1 (lower) subsites, as well as an appendix cavity to the right of S_1. Carbon atoms are colored yellow, oxygen atoms red, and nitrogen atoms blue.

Table 3: Specific Amino Acid Replacements Thought To Be Involved in Change in Substrate Specificity from Hydrophobic (CPD-Y) to Basic (CPD-WII)

CPD-Y	CPD-WII	consequence (CPD-Y > CPD-WII)
S1′ subsite		
Phe 64	Glu 64	hydrophobic > acidic
Leu 272	Glu 272	hydrophobic > acidic
Met 398	Glu 398	hydrophobic > acidic
Tyr 269	H_2O	hydrophobic > ordered H_2O
S1 subsite		
Leu 245[a]	Thr 226, Thr 230	hydrophobic > polar
Pro 181	Asp 181	hydrophobic > acidic
Gln 184	Asp 184	polar > acidic
beyond the S1 subsite		
Phe 300	Asp 303B	hydrophobic > acidic
Arg 304	Thr 307, 307 O	hydrophobic (Cβ, CG, CD) > polar
Leu 307/Phe 308	deletion	hydrophobic > occupied by bulk solvent near α-helix C-terminus
Val 230,[a] Trp 231,[a] Val 234,[a] Pro 235[a]	deletion	hydrophobic > occupied by bulk solvent
Ile 238[a]	Asp 223, 218 O, 219 O, H_2O 101	hydrophobic > acidic, polar
Asn 241[a]	Asp 223, H_2O 94, H_2O 193	polar > acidic, polar
Ala 246[a]	Asp 227	hydrophobic > acidic
Trp 312	bulk solvent, 215 O, 216 O	hydrophobic > polar

[a] Helix 230–251 is translated 5 Å toward the CPD-Y active site relative to CPD-WII.

240, and 224–233, such that the majority of the amino acids corresponding to the hydrophobic patch of CPD-Y can be identified. CPD-MIII also appears to have an extensive solvent-exposed hydrophobic patch at the edge of its active site, while in CPD-S1, this patch is predicted to be acidic and polar. An interesting possibility is that the hydrophobic surface

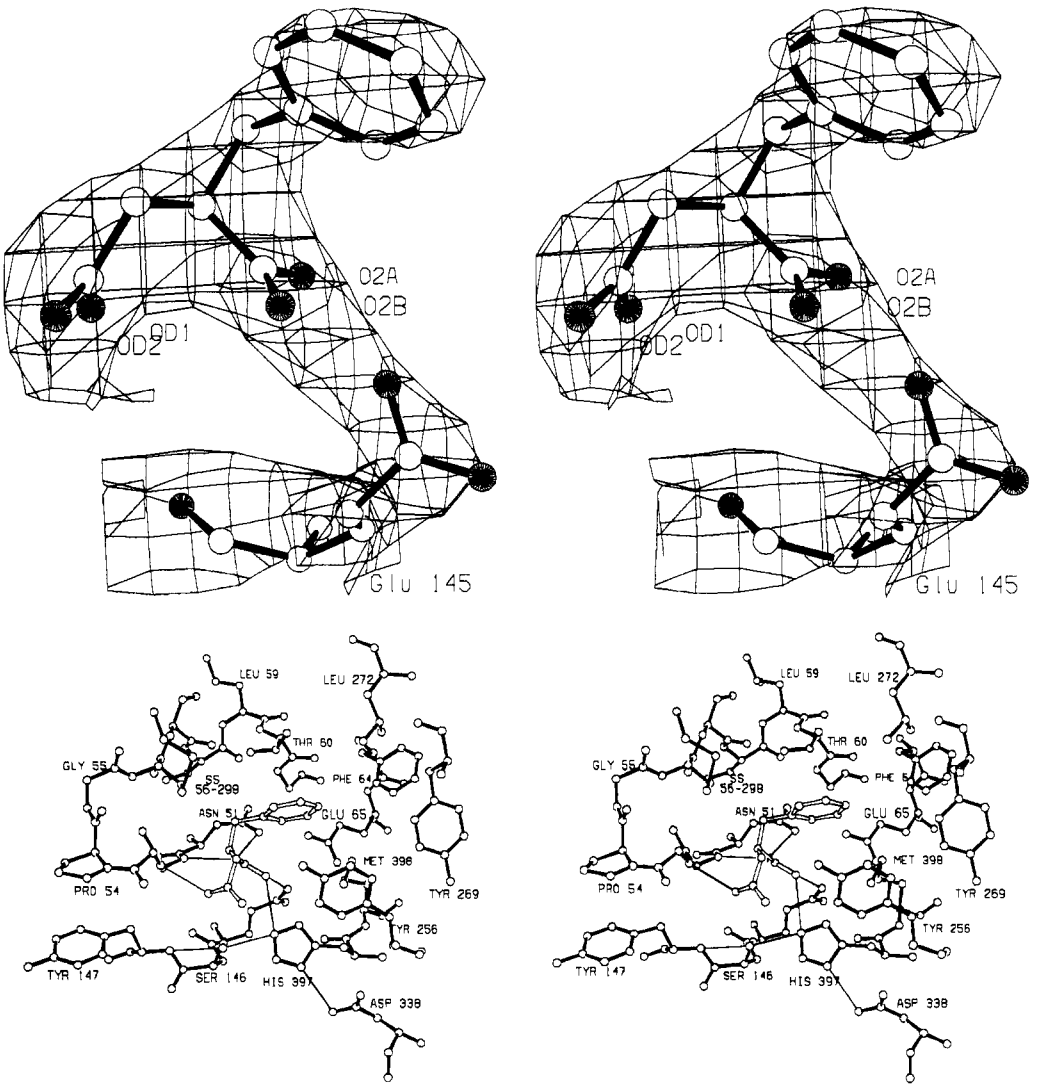

FIGURE 8: (a, top) Final 3.2-Å $2F_o - F_c$ density contoured at 1 SD for "biproduct analog" BZS and Glu 145, as seen in the complex with CPD-Y. (b, bottom) BZS binding site of CPD-Y. Potential hydrogen bonds (between 2.5 and 3.5 Å) are indicated by thin lines. Hydrogen bonds between atom OD2 of BZS, which has no analogous atom in substrate, and CPD-Y are omitted for clarity.

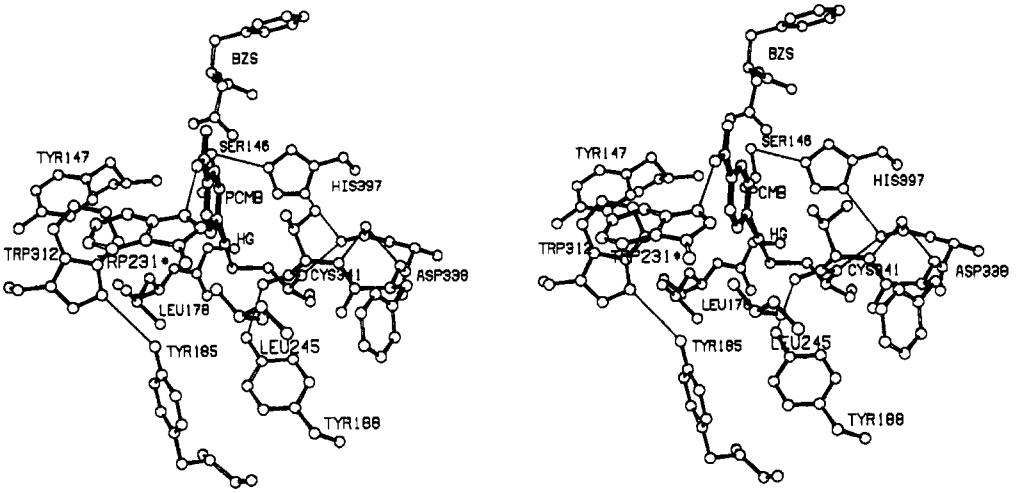

FIGURE 9: Final model of the CPD-Y/pCMB covalent adduct including the catalytic triad and surrounding residues. Note the hydrogen bond between the pCMB carboxylate and Trp 231 of a symmetry related molecule. The location of symmetry-related Trp 231 indicates the position of a potential subsite for binding side chains of peptide substrates.

in CPD-Y assists in the hydrolysis of hydrophobic peptides, which may adopt collapsed conformations with the C-terminal peptide bond inaccessible to cleavage. An extended hydrophobic surface could help unfold collapsed hydrophobic peptides, making the C-terminus accessible for hydrolysis.

Disulfide Bridges. Disulfide bonds occur between residues 56–303, 210–222, and 246–268 of CPD-WII and residues 56–298, 193–207, 217–240, 224–233, and 262–268 of CPD-Y. Disulfide 56–298, which is conserved in known serine carboxypeptidase sequences, is located at the edge of the active

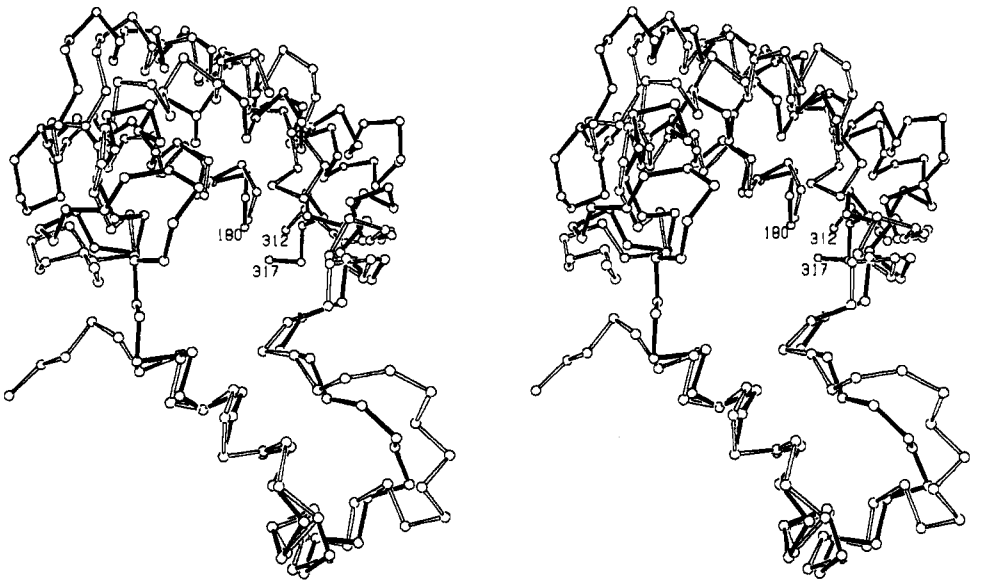

FIGURE 10: Comparison of the "insertion" domains of CPD-Y (dark bonds) and CPD-WII (light bonds), residues 180–317 and 180–312, respectively.

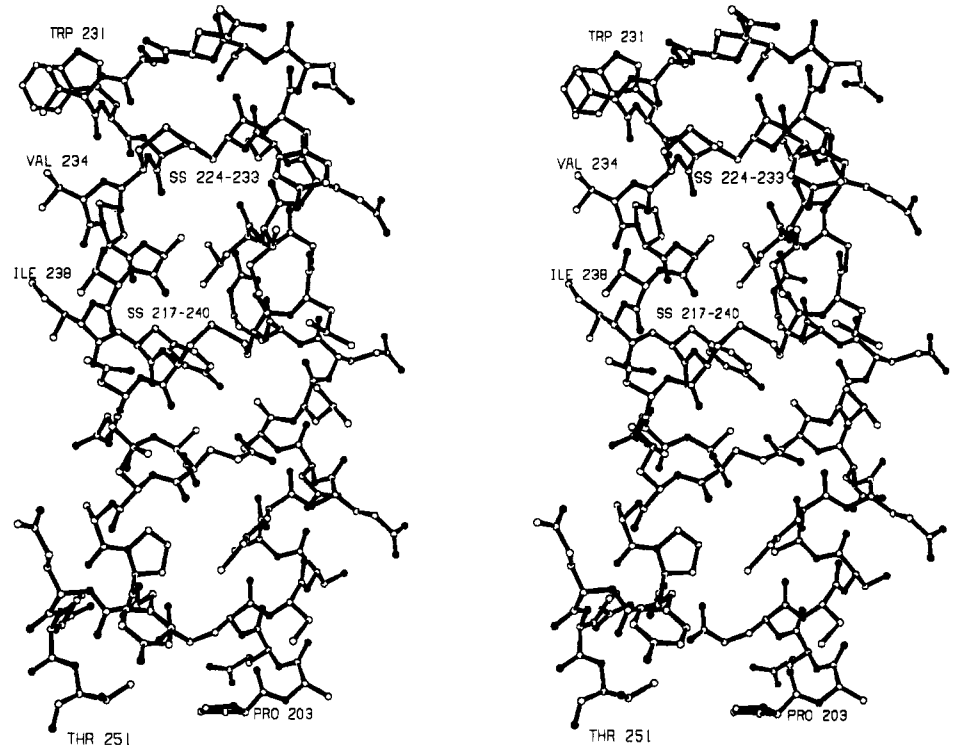

FIGURE 11: Novel "disulfide zipper", consisting of disulfides 217–240 and 224–233, covalently linking antiparallel α-helices 204–227 and 230–251. This motif appears to be conserved between CPD-Y, CPD-S1, and CPD-MIII (see text).

site close to the serine 146 nucleophile and is probably important for catalysis, for example, by stabilizing the distorted main chain of active site residues Gly 52 and Gly 53, which directly interact with substrate (see below).

Superposition of backbone structures on the basis of the 255 "core" α-carbons and by comparing distances between matched disulfides suggests that three of them are spatially conserved: 56–303 (CPD-WII) and 56–298 (CPD-Y), 210–222 and 217–240, and 246–268 and 262–268. Disulfide 217–240, although translated 3 Å relative to 210–222 of CPD-WII, links equivalent turns of antiparallel α-helices; thus, they are topologically equivalent. Disulfide 262–268 is translated 9 Å relative to 246–268 of CPD-WII, possibly for functional reasons (see below).

These matched disulfides are located on one face of the molecule in a triangular arrangement surrounding the active site. However, the yeast enzyme has additional disulfides between residues 193–207 and 224–233, located on the same face of the molecule, resulting in a ring of five disulfide bonds surrounding the perimeter of the active site cavity (Figure 3). This arrangement of disulfides suggests that they play an important role in stabilization of the extensive active site cavity of serine carboxypeptidases.

Potential Role of the CPD-Y Pro Sequence in Folding. Upon synthesis and transport into the endoplasmic reticulum, proCPD-Y contains an additional 91 N-terminal amino acids that have been shown to contain the vacuolar sorting determinant (Valls et al., 1990). ProCPD-Y is also inactive,

so the pro sequence has at least two functions. Winther and Sørensen (1991) have shown that proCPD-Y can be reversibly denatured but mature CPD-Y cannot and that the pro sequence, predicted to be largely α-helical, is required for proper folding. The isolated CPD-Y pro sequence was found to have a relatively high content of regular secondary structure but little tertiary structure (Sørensen et al., 1993). They propose that this partially folded character is important for the function of the CPD-Y pro region as a "cotranslational chaperone".

As discussed earlier, the active site of CPD-Y contains a large exposed hydrophobic patch which does lead to aggregation (crystallization) and would be unlikely to spontaneously form during renaturation of mature CPD-Y. The pro sequence of CPD-Y may prevent aggregation of incompletely folded intermediates by functioning as a cotranslational chaperone as proposed by Ellis (1990), protecting this hydrophobic patch until folding is complete. We suggest that the extra disulfides present in CPD-Y versus CPD-WII may be important to stabilize the final mature form of the enzyme. The latter idea is supported by the observation that CPD-Y requires protein disulfide isomerase (PDI) to form wild-type disulfide bonds (LaMantia & Lennarz, 1993). In mutant strains lacking disulfide isomerase activity, proCPD-Y containing non-native disulfides accumulates in the endoplasmic reticulum.

Similar results have been described by Baker and Agaard (1992), who showed that when mature α-lytic protease is denatured and then renatured, it becomes kinetically trapped in an inactive misfolded state. This inactive form can be converted to an active conformation upon addition of the 167 amino acid pro sequence.

Insertions and Deletions and Changes in Secondary Structure. Relative to CPD-WII, CPD-Y has 8 insertions totaling 33 amino acids and 9 deletions totaling 31 amino acids (see Figure 4). In addition to lengthening of the helices in the "insertion domain", other insertions and deletions may be related to the difference in substrate specificity between CPD-Y and CPD-WII. A deletion of one amino acid at the N-terminal end of α-helix 59–62 shortens it by one residue, altering one side of the S_1' subsite (see below). A deletion of five amino acids occurs between Cys 262 and Cys 268 and may be partly responsible for the single-chain state of CPD-Y versus the two-chain CPD-WII (the shorter disulfide-bridged loop of CPD-Y may be more resistant to proteolysis). Residues 260–271 of CPD-Y follow a path much closer to the active site than their CPD-WII counterparts, and Tyr 269 of CPD-Y replaces several well-ordered water molecules in the CPD-WII structure (see below). Deletion of three amino acids after Ser 106 results in a structural transition from the 3_{10} helix (residues 107–112 of CPD-WII) to an irregular strand that skirts the previous location of the 3_{10} helix, suggesting a possible binding site for extended substrates (see below). An insertion of five amino acids at the N-terminus of β-strand 10 is associated with a new α-helix in CPD-Y, comprised of residues 358–363. Other insertions and deletions and minor structural rearrangements not likely to be related to specificity or activity occur in surface loops and are summarized in Figure 4.

Comparison of Catalytic Residues. Both CPD-Y and CPD-WII contain a Ser 146-Asp 338-His 397 catalytic triad similar to those of subtilisin and trypsin-like serine endoproteases. However, residues comprising the catalytic triad are positioned by a unique fold and are in a different order in the linear sequence in all three families.

Atoms comprising the catalytic triad of CPD-Y and CPD-WII (including the amide nitrogens of the oxyanion hole) were compared and found to superimpose with an rms coordiate error of 0.29 Å, which is at the level of the coordinate error of the two structure determinations. They are therefore in identical configurations for the purposes of discussion. Considering that molecular replacement was not used in this instance and that each structure was determined independently by manual model building and crystallographic refinement, this is a remarkable level of agreement.

A number of unusual features noted for CPD-WII (Liao et al., 1992) are also present in the active site of CPD-Y. The side chain and main chain of the Ser 146 nucleophile are in strained conformations in a tight turn between the N-terminus of an α-helix and the C-terminus of a strand of β-sheet. The main-chain torsional angles of Ser 146 of CPD-Y are in a "disallowed" region of the Ramachandran diagram at $(\Phi,\Psi) = (62°, -107°)$, which this protein fold requires for maximum exposure of the short Ser side chain to incoming substrate (Ollis et al., 1992). Atom Ser 146 OG is near hydrogen-bonding distance to His 397 NE2 in both structures (3.7 Å in CPD-Y, 3.3 Å in CPD-WII). The largest difference between respective catalytic triads is the location of Ser 146 OG, which has high thermal parameters in both structures, indicating rotational flexibility about the Cα–Cβ bond.

Histidine 397 makes a salt bridge with Asp 338, the latter of which is totally solvent inaccessible. As discussed by Liao et al. (1992) for CPD-WII, the plane of the carboxylate is about 45° from parallel to the plane of the imidazole, a geometry far from optimal for proton transfer. This feature distinguishes the catalytic triad of these enzymes from those of the trypsin and subtilisin families. Finally, all but 8 Å^2 of His 397 is shielded from solvent by a perpendicular interaction between the imidazole ring of His 397 and Tyr 256 of CPD-Y (Tyr 239 of CPD-WII). The apparent conservation of Tyr 256 in all known serine carboxypeptidase sequences and its interaction with His 397 suggest that it is functionally important.

Residues Involved in Substrate Recognition. There are at least three distinct binding subsites in serine carboxypeptidases; the terminal carboxylate binding site, the S_1' subsite [which binds the P_1' side chain of the substrate, notation of Schecter and Berger (1967)], and the S_1 subsite. Binding sites beyond S_1 have not been identified, but the enzyme has low activity toward dipeptide substrates, suggesting that residues beyond P_1 are important for formation of the Michaelis complex.

(i) Carboxylate Binding Site. The C-terminal binding site for a peptide substrate was investigated by studying the binding of the "bi-product inhibitor" benzylsuccinic acid (BZS) (Byers & Wolfenden, 1973) to CPD-Y at pH 4.5, the optimal pH for peptidase activity (Breddam, 1986). The mode of binding (see Figure 8) was compared to that of arginine to CPD-WII (Liao et al., 1992). Residues that bind the C-terminal carboxylate in both complexes, Asn 51, Gly 52, and Glu 145, are conserved in all known sequences and superimpose to within coordinate error between CPD-Y and CPD-WII (except for the side chain of Glu 145, which is not well localized in CPD-Y). Furthermore, residues that orient these side chains for interaction with C-terminal carboxylate, Trp 49 and Glu 65, are also conserved in sequence and position. The BZS α-carboxylate, which mimics the C-terminal carboxylate of a peptide substrate, forms hydrogen bonds with Asn 51 ND2, the amide of Gly 52, Glu 145 OE1, and His 397 NE2. The α-carboxylate does not occupy precisely the same position on CPD-Y as does the carboxylate of arginine on CPD-WII;

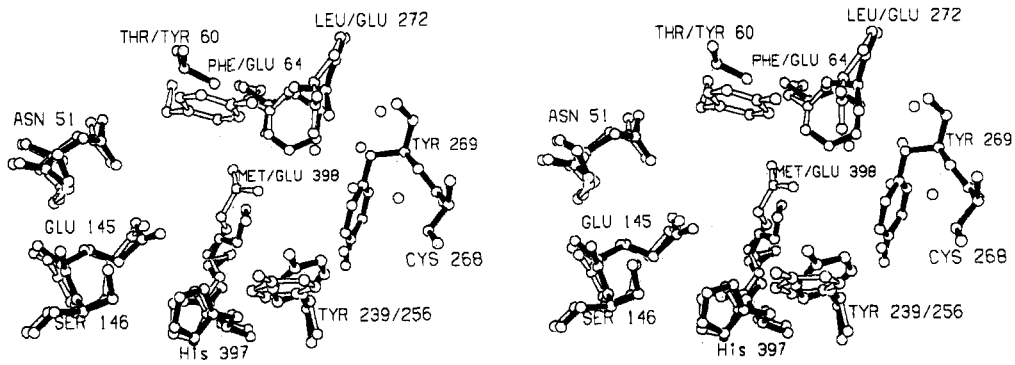

FIGURE 12: Superposition of putative S_1' subsites of CPD-Y (dark bonds) and CPD-WII (light bonds) based on a conserved core of 255 α-carbons. Hydrophobic residues Phe 64, Leu 272, and Met 398 of CPD-Y replace Glu 64, Glu 272, and Glu 398 of CPDW-II. The water molecules near Tyr 269 are seen only in the CPD-WII S_1' subsite.

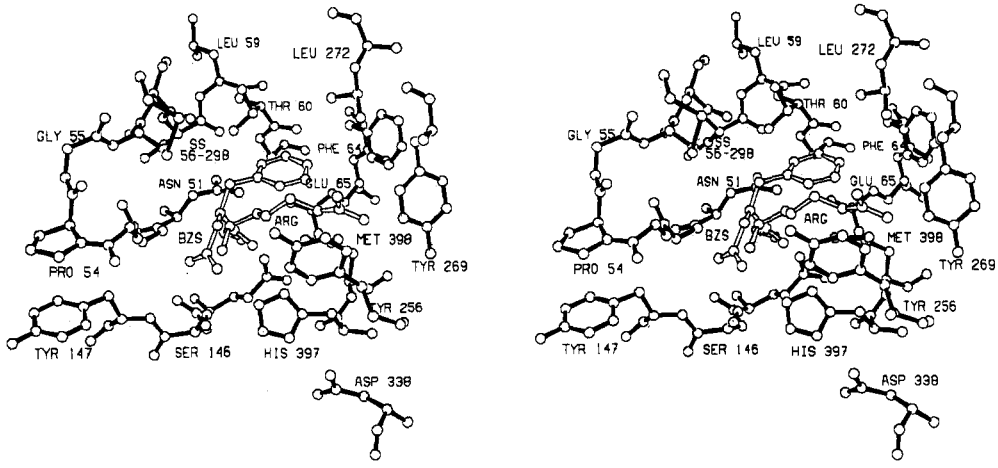

FIGURE 13: Comparison of a reaction product arginine bound to CPD-WII with BZS bound to CPD-Y based on an alignment of active site residues (see text). The arginine side chain would sterically overlap with Tyr 269 if bound to CPD-Y in the same orientation as to CPD-WII.

rather the binding to the oxyanion hole as well requires some adjustments at this site. In the final model, His 397 NE2 is within hydrogen-bonding distance to Ser 146 OG (2.8 Å) and the BZS α-carboxylate (3.4 Å) (Figure 8b). This close proximity raises the possibility that binding of the C-terminal carboxylate of peptide substrates might have a role in the enzymatic mechanism, as is discussed in detail in the accompanying paper (Bullock et al., 1994). They suggest that the carboxylate of incoming substrate could accept a proton from His 397, resulting in "substrate-assisted" activation of the serine nucleophile. This term was coined by Carter and Wells (1987) in reference to their experiments in which an engineered subtilisin was made specific for substrates with histidine at P_2. If this suggestion is borne out, serine carboxypeptidases may represent the first example of substrate-assisted catalysis to be observed in nature.

The BZS β-carboxylate is within hydrogen-bonding distance to His 397 and the amide of Gly 53 and is near the amide of Tyr 147 (3.9 Å), demonstrating that this site is well suited to binding a negative charge as in the proposed tetrahedral intermediates of the hydrolysis reaction.

(ii) S_1' Binding Site. The phenyl moiety of BZS occupies the putative S_1' subsite, which is largely hydrophobic and is comprised of Thr 60, Phe 64, Glu 65, Tyr 256, Tyr 269, Leu 272, Met 398, and disulfide 56–298 (Figure 8b). In CPD-Y, Phe 64, Leu 272, and Met 398 replace Glu 64, Glu 272, and Glu 398 of CPD-WII, reiterating the theme that replacement of polar groups with hydrophobics correlates with the change in substrate specificity from basic to hydrophobic. However, other structural changes may be related to the difference in specificity. Residues 260–271 of CPD-Y adopt completely different conformations in the two structures, and disulfide 262–268 anchors Tyr 269 in a position occupied by several ordered water molecules in CPD-WII (water molecules at this site are involved in binding arginine at the S_1' of CPD-WII). A deletion of one amino acid shortens α-helix 59–62 by one amino acid relative to CPD-WII. Combined with replacement of Tyr 60 of CPD-WII by Thr 60, the S_1' subsite of CPD-Y is slightly "deeper", which may assist in substrate discrimination. An overlay of putative S_1' subsites is shown in Figure 12. With the exception of Met 398 SD, there are no available hydrogen bond donors or acceptors within 4 Å of the BZS phenyl ring. The location of Met 398 in the S_1' subsite is consistent with studies involving modification of Met 398 by chemical means (Breddam & Svendsen, 1984) and mutagenesis (Bech et al., 1985; Winther et al., 1985), both of which alter CPD-Y substrate preference at S_1'. The observation that Thr 60 and Met 398 move in response to BZS binding indicates a degree of "plasticity" at S_1' that may be necessary in a subsite capable of binding several different amino acids. An overlay of the CPD-Y/BZS complex and the CPD-WII/Arg complex is shown in Figure 13. A comparison of the binding of BZS to CPD-Y, to CPD-WII, and also to carboxypeptidase A (Mangani, 1992) is discussed elsewhere.

(iii) S_1 Binding Site. Previous studies involving modifications at Cys 341 had suggested that it be near or part of the S_1 subsite (Breddam & Kanstrup, 1987; Winther & Breddam, 1987). Stoichiometric modification of Cys 341 by *p*-(chloromercuri)benzoic acid (pCMB) alters CPD-Y substrate preference at P_1 toward smaller side chains (Bai & Hayashi, 1979, Breddam, 1983). A model was constructed and refined

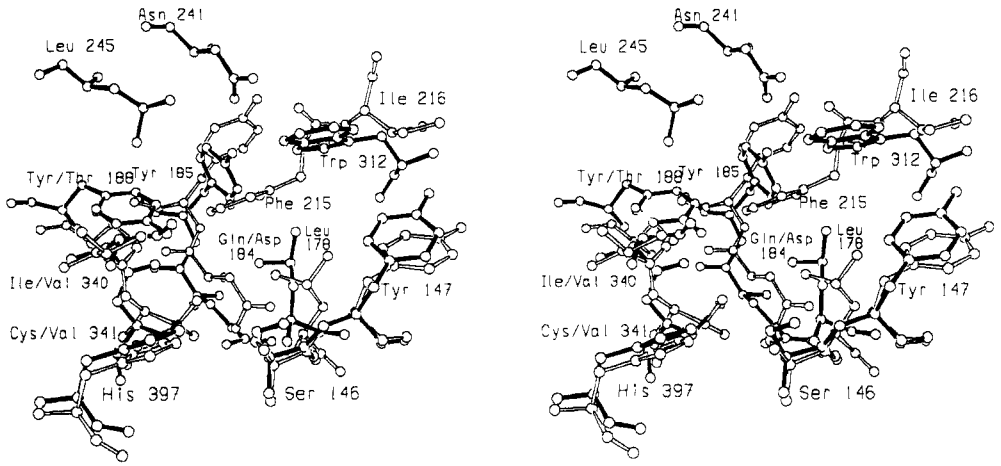

FIGURE 14: Superposition of putative S_1 subsites of CPD-Y (dark bonds) and CPD-WII (light bonds) based on the conserved core α-carbons. Although both S_1 subsites are primarily hydrophobic, α-helix 230–251 moves toward the S_1 subsite of CPD-Y, contributing Asn 241 and Leu 245 where there is only bulk solvent in the CPD-WII model. In addition, structural rearrangements place Trp 312 of CPD-Y at the edge of the S_1 subsite, in a position occupied by Phe 215 and Ile 216 of CPD-WII. Residues Asp 181 and Asp 184 coordinate several water molecules near the bottom of the putative S_1 subsite of CPD-WII.

against data collected from pCMB-modified CPD-Y. In the final model, pCMB lies in a hydrophobic pocket opposite the Ser-His diad from the S_1' subsite, surrounded by Tyr 147, Leu 178, Tyr 185, Tyr 188, Leu 245, Trp 312, Ile 340, and Cys 341 (Figure 9). Residues Tyr 185 and Tyr 188 form an edge–face aromatic interaction, and the α-helix becomes a 3_{10} helix for one turn. Before the structure of CPD-Y was known, the amino acid sequence was aligned with that of CPD-WII (the only available three-dimensional model of a serine carboxypeptidase) and used to predict residues in the S_1 subsite of CPD-Y. Residues predicted to be in the S_1 subsite of CPD-Y were subjected to random mutagenesis, followed by screening for altered specificity variants. Consistent with the putative S_1 subsite as identified by pCMB binding, mutations at Ile 340, Cys 341, and Leu 178 greatly influenced specificity at S_1, but mutations at Glu 215 and Arg 216 had no effect (Olesen & Kielland-Brandt, 1993). The lengthening of α-helices 204–227 and 230–251 relative to their counterparts in CPD-WII places residues 215 and 216 far from the S_1 subsite (Figure 11). An overlay of putative CPD-Y and CPD-WII S_1 subsites (Figure 14) reveals that Asp 184 and several ordered water molecules lie near the bottom of the putative S_1 subsite of CPD-WII and may be important for the higher activity of CPD-WII toward substrates with basic residues at P_1. Interestingly, an "appendix" cavity exists off to one side of the putative S_1 subsite (Figure 7). This cavity is positioned such that it may represent a degenerate S_1 subsite, but it appears that it could also result in nonproductive peptide binding.

pH Dependence of Activity. Serine carboxypeptidases exhibit maximum carboxypeptidase activity between pH 4.0 and pH 5.5, depending on the enzyme (Breddam, 1986), while trypsin- and subtilisin-like serine endopeptidases are essentially inactive at this pH. As has been demonstrated (Mortensen et al., 1994), the effect of pH on activity toward peptide substrates is almost entirely due to K_m, which increases rapidly above pH 6, an effect that has been attributed to interaction of the C-terminal carboxylate of substrate with Glu 145 (Liao et al., 1992; Mortensen et al., 1994).

The lack of strong dependence of k_{cat} on pH led to the suggestion that the active site His has a very low pK_a (<3; Mortensen et al., 1994). Liao et al. (1992) noted that this may due to Tyr 239, which makes an edge–face interaction with the imidazole and shields His 397 from solvent and could result in a lowered pK_a. In CPD-Y, Tyr 256 is in the same location and also shields all but 8 Å^2 of His 397 from solvent, so this interaction is conserved. Another possibility that should be considered is that His 397 has a normal pK_a and that the enzyme has evolved a means of deprotonating His 397 at the appropriate point in the catalytic cycle, possibly *via* the substrate carboxylate as noted earlier [see accompanying paper by Bullock et al. (1994)].

A striking observation of this study is an unusual interaction in the active site between a pair of glutamic acid residues conserved in all known sequences, Glu 65 and Glu 145. These residues make a hydrogen bond to each other (2.8 Å), and in turn Glu145 is a ligand to the carboxylate of bound BZS (Figure 8b), suggesting that at least two of these groups must be protonated for productive binding to occur. Deprotonation induced by high pH would disrupt peptide binding due to electrostatic repulsion, as has been demonstrated by mutagenesis (Mortensen et al., 1994). That this pair of carboxylates is structurally conserved between two enzymes with such low amino acid sequence homology as CPD-WII and CPY suggests that it has some unusual properties and an important, but not fully understood, role in catalysis.

Comparison with Other α/β Hydrolases. Serine carboxypeptidases share a core fold with several hydrolases whose structures have recently been determined. These include dienelactone hydrolase from *Pseudomonas* sp. B13 (Pathak et al., 1988), haloalkane dehalogenase from *Xanthobacter autotrophicus* (Franken et al., 1991), lipases from *Geotrichum candidum* (Schrag et al., 1991), *Rhizomucor miehei,* and pancreas (Brzozowski et al., 1991), acetylcholinesterase (AChE) from *Torpedo californica* (Sussman et al., 1991), and cutinase from *Fusarium solani* (Martinez et al., 1992). This family of enzymes has a common topology termed the α/β hydrolase fold (Ollis et al., 1992) that anchors a positionally conserved catalytic triad which varies in composition. The only invariant residue in all α/β hydrolase sequences is the histidine of the catalytic triad. Most members of the family have large, mostly α-helical insertions relative to the core fold. The largest insertion domain tends to occur in loops at the C-terminal end of the parallel β-sheet, adopting completely different folding patterns in the various hydrolase-fold enzymes.

In every instance, these insertions have been proposed to play specific roles. For example, AChE has three insertions

relative to the core fold that place 14 aromatic residues around the perimeter of the active site cavity and were suggested to be involved in channeling the substrate to the active site (Tan et al., 1993b). The lipases (Schrag et al., 1991, Brzozowski et al., 1991) have a helical lid, also contributed by the insertion domains, that shields the active site catalytic apparatus from solvent. The lid has been proposed to roll back upon contact with the substrate micelle, making the active site accessible in the activation step. The structures of four homologous fungal lipases reveal a conserved "polar core" near the substrate binding site that may be important for interfacial catalysis (Derewenda et al., 1994), also involving part of the insertion domain.

In this paper, we have suggested that serine carboxypeptidases also have a large insertion relative to the core topology (residues 180–317) that may be functionally important, in this case recognition of peptide substrates. Furthermore, comparison of the amino acid sequence and secondary structure of this region between CPD-Y and CPD-WII suggests that this insertion domain has changed more rapidly during evolution than the rest of the molecule, as there is no statistically significant sequence homology in this region. However, unlike other members of the α/β hydrolase-fold family, the secondary structures of the two insertion domains are still recognizably similar.

We suggest that the insertion domain, residues 180–317 of CPD-Y, is a segment of the molecule that has few structural constraints upon its amino acid sequence, and as a consequence, it can evolve much more rapidly than the rest of the molecule to give rise to new functions. It may be that the "α/β hydrolase core" of the molecule is important for proper folding and placement of catalytic residues and the additions are of lesser importance. This hypothesis helps explain the lack of obvious structural motifs in these insertions in other members of the hydrolase-fold family. These molecules have simply diverged so far that the extra domains have "evolved away" all vestiges of structural similarity, except for their predominantly α-helical character. While one might also consider "intron swaps" as possibly resulting in these rearrangements, the intron structures of a serine carboxypeptidase gene from wheat (although not coding for CPD-WII; Baulcombe et al., 1987) and rice (Washio & Ishikawa, 1992) do not support this idea. Both genes contain eight introns, dispersed fairly uniformly throughout the coding sequence. Intron boundaries are not located at or very close to the places in the fold where the insertion domain can be regarded to be inserted into the basic fold, and there are two different introns in the region coding for the insertion domain. Finally, the gene for CPD-Y contains no introns (Stevens et al., 1986).

In conclusion, the difference in size between members of the α/β hydrolase fold can be rationalized, at least in part, as being due to large insertions relative to a common core that confer specific functionalities. CPD-Y and CPD-WII form an interesting and unusual example, discovered early in the course of divergent evolution, showing that parts of a molecule evolve at different rates than other parts.

ACKNOWLEDGMENT

We thank Drs. Enoch Baldwin, Bob Dubose, Larry Weaver, Dale Tronrud, and Brian Matthews for helpful discussions and technical advice and Karen Kallio for expert technical assistance.

REFERENCES

Bai & Hayashi, R. (1979) *J. Biol. Chem. 254*, 8473–8479.
Baker, D., Sohl, J. L., & Agard, A. (1992) *Nature 356*, 263–265.
Ballou, L., Hernandez, L. M., Alvarado, E., & Ballou, C. E., (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 3368–3372.
Baulcombe, D. C., Barker, R. F., & Jarvis, M. G. (1986) *J. Biol. Chem. 262*, 13726–13734.
Bech, L. M., & Breddam, K. (1989) *Carlsberg Res. Commun. 54*, 165–171.
Bech, L. M., Nielsen, J., Winther, J. R., Kielland-Brandt, M. C., & Breddam, K. (1985) *Carlsberg Res. Commun. 51*, 459–465.
Blow, D. M., Birkoft, J. J., & Hartley, B. S. (1969) *Nature 21*, 337–340.
Breddam, K. (1983) *Carlsberg Res. Commun. 48*, 9–19.
Breddam, K. (1986) *Carlsberg Res. Commun. 51*, 83–128.
Breddam, K., & Svendsen, I. (1984) *Carlsberg Res. Commun. 49*, 639–645.
Breddam, K., & Kanstrup, A. (1987) *Carlsberg Res. Commun. 52*, 65–71.
Brzozowski, A. M., Derewenda, Z. S., Dodson, C. G., Lawson, D. M., Turkenburg, J. P., Bjorkling, F., Huge-Jensen, B., Patkar, S. A., & Thim, L. (1991) *Nature 351*, 491–494.
Bullock, T. L., Branchaud, B., & Remington, S. J. (1994) *Biochemistry* (third of three papers in this issue).
Byers, L. D., & Wolfenden, R. (1973) *Biochemistry 12*, 2070–2078.
Carter, P., & Wells, J. A. (1987) *Science 237*, 394–399.
Cleland, W. W. (1992) *Biochemistry 31*, 317–319.
Connolly, M. L. (1983) *Science 221*, 709.
Cooper, A., & Bussey, H. (1989) *Mol. Cell. Biol. 9*, 2706–2714.
Derewenda, U., Swenson, L., Green, R., Wei, Y., Dodson, G. G., Yamaguchi, S., Haas, M. J., & Derewenda, Z. S. (1994) *Nature Struct. Biol. 1*, 36–47.
Ellis, R. J. (1990) *Semin. Cell. Biol. 1, 1–9.*
Franken, S. M., Rozeboom, H. J., Kalk, K. H., & Dijkstra, B. W. (1991) *EMBO J. 10*, 1297–1302.
Hayashi, R., Moore, S., & Stein, W. H. (1973) *J. Biol. Chem. 248*, 8366–8369.
Hayashi, R., Bai., Y., & Hata, T. (1975) *J. Biol. Chem. 250*, 5221–5226.
Howard, A. J., Nielsen, C., & Xuong, N. H. (1985) *Methods Enzymol. 114*, 452–471.
Johansen, J. T., Breddam, K., & Ottesen, M. (1976) *Carlsberg Res. Commun. 41*, 1–14.
Jones, T. A., & Thurip, S. (1986) *EMBO J. 5*, 819–822.
Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577–2637.
Kossiakoff, A. A., & Spencer, S. A. (1981) *Biochemistry 20*, 6462–6474.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946–950.
LaMantia, M., & Lennarz, W. J. (1993) *Cell 74*, 899–908.
Lesk, A. M., & Chothia, C. (1980) *J. Mol. Biol. 136*, 225–270.
Liao, D.-I., & Remington, S. J. (1990) *J. Biol. Chem. 265*, 6528–6531.
Liao, D.-I., Breddam, K., Sweet, R. M., Bullock, T., & Remington, S. J. (1992) *Biochemistry 31*, 9796–9812.
Lijas, A., & Rossman, M. G. (1974) *Annu. Rev. Biochem. 43*, 475.
Mangani, S. (1992) *J. Mol. Biol. 223*, 573–578.
Martinez, C., De Geus, P., Lauwereys, M., Matthyssens, G., & Cambillau, C. (1992) *Nature 356*, 615–618.
Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491–497.
Matthews, B. W., Sigler, P. B., Hendersen, R., & Blow, D. M. (1967) *Nature 214*, 652.
Mattos, C., Rasmussen, B., Ding, X., Petsko, G. A., & Ringe, D. (1994) *Nature Struct. Biol. 1*, 55–58.
Mortensen, U. H., Remington, S. J., & Breddam, K. (1994) *Biochemistry 33*, 508–517.
Odya, C. E., & Erdös, E. G. (1981) *Methods Enzymol. 80*, 460–466.
Olesen, K., & Kielland-Brandt, M. C. (1993) *Protein Eng. 6*, 409–415.

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. 5*, 197–211.

Pathak, D., Ngai, K. L., & Ollis, D. (1988) *J. Mol. Biol. 204*, 435–445.

Remington, S. J., & Breddam, K. (1994) *Methods Enzymol.* (in press).

Rossmann, M. G., & Argos, P. (1976) *J. Mol. Biol. 105*, 75–95.

Sadis, S., Raghavendra, K., & Hightower, L. E. (1990) *Biochemistry 29*, 8199–8206.

Schechter, I., & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157–167.

Schrag, J. D., Li, Y., Wu, S., & Cygler, M. (1991) *Nature 351*, 761–764.

Sørensen, S. B., Svendsen, I., & Breddam, K. (1987) *Carlsberg Res. Commun. 52*, 285–296.

Sørensen, P., Winther, J. R., Kaarsholm, N. C., & Poulsen, F. M. (1993) *Biochemistry 32*, 12160–12166.

Steigemann, W. (1974) Doctoral Thesis, Technical University, Munich.

Stevens, T. H., Rothman, J. H., Paynes, G. S., & Schekman, R. (1986) *J. Cell Biol. 102*, 1551–1554.

Sussman, J. L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L., & Silman, I. (1991) *Science 253*, 872–879.

Svendsen, I., Hofmann, T., Endrizzi, J., Remington, S. J., & Breddam, K. (1993) *FEBS Lett. 333*, 39–43.

Tan, F., Morris, P. W., & Skidgel, R. A. (1993a) *J. Biol. Chem. 268*, 16631–16638.

Tan, R. C., Truong, T. N., McCammon, J. A., & Sussman, J. L. (1993b) *Biochemistry 32*, 401–403.

Tronrud, D. E. (1992) *Acta Crystallogr. A48*, 912–916.

Tronrud, D. E., TenEyck, L. F., & Matthews, B. W. (1987) *Acta Crystallogr. A43*, 489–501.

Valls, L. A., Hunter, C. P., Rothman, J. H., & Stevens, T. H. (1987) *Cell 48*, 887–897.

Valls, L. A., Winther, J. R., & Stevens, T. H. (1990) *J. Cell Biol. 111*, 361–368.

Washio, K., & Ishikawa, K. (1992) *Plant Mol. Biol. 19*, 631–640.

Winther, J. R., & Breddam, K. (1987) *Carlsberg Res. Commun. 52*, 263–273.

Winther, J. R., & Sørensen, P. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9330–9334.

Winther, J. R., Kielland-Brandt, M. C., & Breddam, K. (1985) *Carlsberg Res. Commun. 50*, 273–284.

Winther, J. R., Stevens, T. H., & Kielland-Brandt, M. C. (1991) *Eur. J. Biochem. 197*, 681–689.