Crystal Structure of Kex1Δp, a Prohormone-Processing Carboxypeptidase from *Saccharomyces cerevisiae*[†,‡]

Brian H. Shilton,[§] David Y. Thomas,[||] and Miroslaw Cygler*[,§]

Biotechnology Research Institute, National Research Council, 6100 Royalmount Avenue, Montréal, Québec H4P 2R2, Canada

Received February 26, 1997[⊗]

ABSTRACT: Kex1p is a prohormone-processing serine carboxypeptidase found in *Saccharomyces cerevisiae*. In contrast to yeast serine carboxypeptidase (CPD-Y) and wheat serine carboxypeptidase II (CPDW-II), Kex1p displays a very narrow specificity for lysyl or arginyl residues at the C-terminus of the substrate. The structure of Kex1Δp, an enzyme that lacks the acidic domain and membrane-spanning portion of Kex1p, has been solved by a combination of molecular replacement and multiple isomorphous replacement and refined to a resolution of 2.4 Å. The S1′ site of Kex1Δp is sterically restricted compared to those from CPD-Y or CPDW-II; it also contains two acidic groups that are well positioned to interact with the basic group of a lysine or arginine side chain. The high specificity of Kex1p can therefore be explained by a combination of steric and electronic factors. The structure of the S1 site of Kex1Δp is also well suited for binding of a lysine or arginine side chain, and the enzyme may therefore exhibit a preference for these residues at P1.

The yeast *Saccharomyces cerevisiae* shares with other eukaryotes a proteolytic system for the posttranslational processing of precursors of secreted protein and peptide hormones. The Kex (for *K*iller *ex*pression) prohormone processing system produces mature α-pheromone (Dmochowska *et al.*, 1987; Zhu *et al.*, 1987) and active K1 killer toxin (Wickner & Leibowitz, 1976). The Kex system processes prohormones in two steps: in the first step, Kex2p, a subtilisin-like endoprotease, catalyzes cleavage after a sequence of two basic residues (lysine or arginine). In the second step, Kex1p,[1] a serine carboxypeptidase, catalyzes removal of the C-terminal lysyl and/or arginyl residues. Both Kex enzymes are membrane-anchored and reside in a *trans*-Golgi compartment (Bryant & Boyd, 1993; Fuller *et al.*, 1989).

The processing steps carried out in yeast are the same as those found in mammalian cells (Hutton, 1990; Skidgel, 1988). Genes coding for a family of Kex2p-like convertases—furin and PC1−PC7—have been identified in human, mouse, and other metazoans and have been shown to cut after a sequence of two basic residues [for review see Denault and Leduc (1996) and Seidah *et al.* (1994)]. The carboxypeptidases thought to be involved in mammalian prohormone processing are often metalloproteases (Skidgel, 1988); however, two mammalian serine carboxypeptidases that are related to Kex1p have been identified: human protective protein (HPP or Cathepsin A; Galjart *et al.*, 1988, 1991) and human prolylcarboxypeptidase (Tan *et al.*, 1993). The correspondence of the Kex1−Kex2 processing system was demonstrated when the genes for these enzymes were transferred to mammalian neuroendocrine cells and shown to process correctly proopiomelanocortin (Germain *et al.*, 1990; Thomas *et al.*, 1990; Zollinger *et al.*, 1990). Thus, in both mammalian and yeast systems the endoproteases, which are related by their sequence homology, are very similar in their substrate preferences, and the carboxypeptidases share a specificity for C-terminal lysine and arginine residues.

Kex1p bears sequence homology (approximately 25% in the catalytic domain) to other serine carboxypeptidases, including three for which the crystal structures have been determined: wheat serine carboxypeptidase II (CPDW-II; Liao *et al.*, 1992; Liao & Remington, 1990), yeast serine carboxypeptidase (CPD-Y; Endrizzi *et al.*, 1994), and human protective protein (HPP; Rudenko *et al.*, 1995). The serine carboxypeptidases belong to the α/β-hydrolase family of enzymes (Ollis *et al.*, 1992) and are therefore related to the lipases and esterases, which are of long-standing interest in this laboratory (Cygler *et al.*, 1993).

It is interesting to compare catalysis by the serine carboxypeptidases with that by the serine endoproteases. Stabilization of the enzyme−substrate transition state, rather than the ground state, is a crucial feature of the catalytic mechanism of the serine endoproteases (Fersht, 1988). The heart of the serine endoprotease catalytic machinery, namely, the catalytic triad and oxyanion hole, has evolved independently in the serine carboxypeptidases, with an intriguing difference: the arrangement of these elements is the mirror image of that found in the serine endoproteases. Thus, the tetrahedral transition state formed during catalysis by the serine carboxypeptidases will be enantiomeric to that in the serine endopeptidases. There are, additionally, two special demands placed on these enzymes. The first is that the serine

[†] NRCC Publication 39969. This work was done as part of the official government duties.

[‡] Coordinates of kex1Δp have been deposited with the Brookhaven Protein Data Bank under filename 1AC5.

* Author to whom correspondence should be addressed: tel (514) 496-6321; fax (514) 496-5143; email mirek.cygler@bri.nrc.ca.

[§] Also affiliated with the Montréal Joint Centre for Structural Biology.

[||] Also affiliated with the Biology Department and Department of Anatomy and Cell Biology, McGill University.

[⊗] Abstract published in *Advance ACS Abstracts,* June 15, 1997.

[1] Abbreviations: CPD-Y, yeast serine carboxypeptidase; CPDW-II, wheat serine carboxypeptidase II; Kex1p, killer expression yeast serine carboxypeptidase; Kex1Δp, truncated soluble form of Kex1p; HPP, human protective protein (otherwise known as cathepsin A); PEG-MME, poly(ethylene glycol) methyl ether; rms, root mean square; Bz-FAR, benzoyl-phenylalanyl-alanyl-arginine; Z, benzyloxycarbonyl.

S0006-2960(97)00433-9 CCC: $14.00 Published 1997 by the American Chemical Society

Table 1: Data Collection and Phasing Statistics

	laboratory source (120 K, $\lambda = 1.5418$ Å)	Photon Factory (Tsukube, Japan) (4 °C, $\lambda = 1.0$ Å)		
	native	native	$HgCl_2$	K_2PtCl_4
cell (*a*, *b*, *c*)	57.15, 83.05, 111.11	56.62, 83.98, 111.76	56.82, 84.12, 112.28	55.99, 83.99, 111.91
resolution (Å)	2.4	2.8	3.2	3.2
no. of unique reflections	20693	12921	7279	7909
redundancy	5.11	4.09	3.60	3.72
% completeness[a,b]	91.9 (81.2)	86.3 (55.5)	73.7 (47.8)	78.8 (51.3)
R_{merge}[b]	5.1 (15.4)	5.8 (23.9)	5.4 (19.0)	5.7 (22.3)
R_{iso}[c]	N/A	N/A	15.2	9.9
phasing power[d] (centric/acentric)	N/A	N/A	1.89/2.02	0.77/0.84
Cullis R[d] (centric/acentric/anomalous)	N/A	N/A	0.48/0.61/0.92	0.76/0.86/N.A.
figure of merit[d] (centric/acentric/overall)		0.7160/0.4957/0.5298		

[a] For reflections with $I > \sigma I$. [b] Values in parentheses correpond to the highest resolution shell. [c] As output by the program SCALEIT (CCP4, 1994). [d] As output by the program MLPHARE (CCP4, 1994).

carboxypeptidases, and Kex1p in particular, must recognize the C-terminal residue; therefore, the P1′ position is important for determining the specificity rather than the P1 position as in the trypsin family. Second, the natural substrates contain a free carboxylate that must bind in the general proximity of the catalytic machinery. From the structures of CPD-Y, CPDW-II, and HPP, as well as kinetics studies (Mortensen *et al.*, 1994), it appears that the C-terminal carboxylate is bound in the vicinity of two conserved glutamic acid residues. Thus, in the enzyme−substrate complex there are three carboxylate groups that are essentially buried and forced into proximity, all of which are involved in transition-state stabilization (Stennicke *et al.*, 1996).

As one might expect for a prohormone-processing carboxypeptidase, Kex1p demonstrates an impressive specificity for its substrates: in the case of α-pheromone-KR, the only product detected after a prolonged *in vitro* incubation with Kex1p was mature α-pheromone (Latchinian-Sadek & Thomas, 1993). The C-terminal sequence of the α-peptide of mature K1 killer toxin provided evidence that, after endoproteolytic cleavage by Kex2p, Kex1p had removed only the two C-terminal arginyl residues from the one subunit of the toxin (Zhu *et al.*, 1987). Peptides with C-terminal lysyl or arginyl residues were able to inhibit cleavage of Bz-FAR by Kex1Δp, but no inhibition was observed with the same peptides terminating in leucine, phenylalanine, or glycine (Cooper & Bussey, 1989). The high specificity of Kex1p sets it apart from CPDW-II, CPD-Y, and HPP. CPDW-II and CPD-Y demonstrate preferences for basic or hydrophobic residues, respectively (Breddam *et al.*, 1987; Hayashi *et al.*, 1975), but are able to cleave most other amino acids from the C-terminus. HPP, while not as thoroughly studied as CPDW-II and CPD-Y, also demonstrates a broad specificity (Jackman *et al.*, 1990; Pshezhetsky *et al.*, 1995). The purpose of this study was to delineate the molecular basis for the high specificity exhibited by Kex1p.

Kex1p is a type 1 membrane protein. To facilitate structural and kinetic studies, the Kex1 gene was engineered with a stop codon before the transmembrane domain to produce a soluble form of the enzyme, Kex1Δp (Latchinian-Sadek & Thomas, 1994). After removal of the signal sequence (which occurs during expression in insect cells), Kex1Δp comprises only the catalytic domain (residues 23−505), but it retains all the enzymatic characteristics of its membrane-bound counterpart (Latchinian-Sadek & Thomas, 1993). This protein has been purified and crystallized (Shilton *et al.*, 1996) and we now present the structure of the enzyme, solved by a combination of molecular replacement and multiple isomorphous replacement methods and refined to a resolution of 2.4 Å.

MATERIALS AND METHODS

Purification, Crystallization, and Heavy Atom Derivatization. Kex1Δp was expressed, purified, and crystallized as previously described (Latchinian-Sadek & Thomas, 1994; Shilton *et al.*, 1996). The crystallization mother liquor consisted of 17% PEG-MME ($M_r = 5000$; Aldrich, Milwaukee, WI), 400 mM ammonium acetate, 10 mM sodium azide, and 5% glycerol, pH 6.5. The mercury derivative was obtained by soaking a crystal in mother liquor containing 0.1 mM $HgCl_2$ for 4 h. The platinum derivative was obtained after a 5 day soak in 1 mM K_2PtCl_4.

Data Collection and Processing. Kex1Δp crystallizes in space group $P2_12_12_1$; data and phasing statistics for native and heavy-atom derivative data sets are given in Table 1. All data were processed using the program DENZO (Otwinowski, 1991) and, in all cases, scaled intensities were converted to amplitudes using the program TRUNCATE (CCP4, 1994). A 2.8 Å native data set and the two derivative data sets were collected at beam line 6A2 at the Photon Factory, Tsukuba, Japan, using a wavelength of 1.0 Å. The programs SCALA, AGROVATA, and FHSCALE (CCP4, 1994) were used for internal scaling, merging, and derivative-to-native scaling of these data. A higher resolution (2.4 Å) data set was collected from a frozen crystal using a Rigaku RU300 generator ($\lambda = 1.5418$ Å) and an Raxis-II area detector; in this case, data were scaled and merged using SCALEPACK (Otwinowski, 1991).

Molecular Replacement Models. A structure-based sequence alignment of CPD-Y (PDB code 1YSC; Endrizzi *et al.*, 1994) and CPDW-II (PDB code 3SC2; Liao *et al.*, 1992) was used to align the Kex1Δp sequence and to delineate regions of probable structural conservation among the three enzymes. Models consisting of this conserved core (residues 15−21, 28−37, 45−58, 62−71, 77−104, 113−130, 140−162, 169−186, 272−286, 301−307, 319−350, and 387−413) of either CPDW-II or CPD-Y, as well as the complete structures, were used for a molecular replacement search with the program AMoRe (Navaza, 1994).

Phase Determination and Model Building. Structure factor calculation, phase combination, and map calculations were carried out with the programs SFALL, SIGMAA, and FFT

Table 2: Molecular Replacement of Kex1Δp: Effect of Search Model[a,b]

	rotation function		translation function				rigid-body fitting	
model	peak height[c]	rank	peak height[c]	correlation coefficient	*R*-factor	rank	correlation coefficient	*R*-factor
CPDW-II core	6.1 (4.7)	1	7.7 (5.4)	21.8 (14.2)	51.7 (54.1)	1	23.9	50.9
CPDW-II	8.2 (4.5)	1	7.7 (5.8)	23.0 (17.8)	51.4 (52.8)	1	23.7	51.0
CPD-Y core	6.8 (4.7)	1	7.0 (4.7)	18.3 (14.2)	53.1 (55.0)	1	20.6	52.6
CPD-Y	4.3 (4.4)	2	5.9 (4.9)	15.5 (11.9)	53.7 (54.5)	1	16.3	53.4
CPD-Y/CPDW-II superposition	9.0 (5.2)	1	9.0 (6.2)	27.1 (19.5)	50.2 (52.5)	1	27.8	49.9

[a] Data from 8 to 4 Å were used for all calculations. [b] Numbers in parentheses represent the value for the most prominent incorrect solution. [c] Peak height is the density observed for a given solution divided by the rms deviation from mean density, as output by AMoRe (Navaza, 1994; CCP4, 1994).

(CCP4, 1994). Heavy atom positions were determined from isomorphous difference maps (with coefficients $F_{PH}-F_P$) using phases from the molecular replacement solution of the CPDW-II core model. Heavy atom parameter refinement and phase calculation were carried out using MLPHARE (CCP4, 1994). Model building and analysis were carried out using the program O (Jones *et al.*, 1990).

The structure of Kex1Δp was solved by a combination of multiple isomorphous replacement and molecular replacement. Models containing only the structurally conserved residues from either CPDW-II or CPD-Y yielded clear molecular replacement solutions (Table 2). For comparison, the molecular replacement solutions using complete models of either CPDW-II or CPD-Y are also given. Molecular replacement using a model consisting of a superposition of the complete CPDW-II and CPD-Y structures yielded the strongest solution (Table 2).

Isomorphous difference electron density maps, calculated with phases from the CPDW-II core model, were used to locate heavy atoms. In agreement with the isomorphous difference Patterson map (Shilton *et al.*, 1996), the mercury atom, positioned at (0.018, 0.992, 0.207), was only 1.03 Å away from the z-axis of the unit cell. When the CPD-Y molecular replacement solution was considered, the mercury atom was in a good position to have reacted with Cys341, which is known to react with mercurials (Breddam & Svendsen, 1984). The sequence alignment between Kex1Δp and CPD-Y, coupled with biochemical data demonstrating the inactivation of Kex1Δp by $HgCl_2$ (Latchinian-Sadek & Thomas, 1994), indicated that Kex1Δp also contained a reactive cysteine residue in a similar location (Cys386), confirming the correctness of the molecular replacement solution.

The inclusion of anomalous scattering data from the mercury derivative, and isomorphous data from a weak platinum derivative, produced a 3.2 Å resolution (MIRAS) electron density map that was barely interpretable. However, phase combination with the CPDW-II core molecular replacement model yielded an interpretable electron density map. The CPDW-II core structure was used as the starting point for model-building. The first step was the replacement of nonconserved CPDW-II side chains with those from Kex1Δp. The model was then extended, using the CPD-Y and CPDW-II structures as guides. The MIRAS electron density map, which had no model bias, was used to validate the positioning of new sections of the model. Once easily interpreted sections of the map had been filled, the new model was used for phase combination with the MIRAS data, and updated maps were calculated. This cycle of model building and phase combination continued until the map showed no further signs of improvement. At this stage, the model consisted of 418 residues (out of 482), and refinement in Xplor (Brünger, 1992b) was initiated.

Refinement. The Kex1Δp model, built using data collected from room-temperature crystals at a synchrotron source, was subjected to rigid-body refinement against a higher resolution (2.4 Å) native data set that had been collected from a frozen crystal using a laboratory source (Table 1). The initial *R*-factor for the model was 47%. After several rounds of rebuilding, manual adjustment, simulated annealing (Brünger *et al.*, 1987), energy minimization, and *B*-factor refinement, the model consisted of 458 residues and had an *R*-factor of 30.6% and a free *R*-factor (Brünger, 1992a) of 36.8%. While the quality of the model seemed good, it was not reflected in the *R*-factors, and this prompted a reexamination of the crystallographic data. We found that the crystals were diffracting anisotropically: intensities along the h-axis of the reciprocal cell decreased more rapidly than in the other two directions (Figure 1). Overall anisotropic *B*-factor refinement (Brünger, 1992b) reduced both the conventional *R*-factor and *R*-free by approximately 4% and allowed the refinement to converge. Inclusion of low-resolution data, coupled with bulk solvent correction, further decreased the *R*-factors and greatly improved the quality of the electron density maps.

Structural Analysis. Sequences were aligned using the program SEQSEE (Wishart *et al.*, 1994). Structure superposition was carried out in O (Jones *et al.*, 1990). Superposition over whole structures was initiated with the CA positions of the catalytic triads, followed by refinement using the lsq_improve option in O with a default cutoff for CA distances of 3.5 Å. Solvent accessible surface areas were calculated with the program SURFACE (CCP4, 1994) using a probe radius of 1.4 Å.

RESULTS

The Kex1Δp Model. The Kex1Δ genetic construct lacks the C-terminal acidic domain, the transmembrane region, and the cytosolic domain of Kex1. Prior to crystallization, N-terminal sequencing of purified Kex1Δp indicated that the signal peptide had been properly cleaved; therefore, Kex1Δp consisted of Leu23−Thr505 of Kex1. The three-dimensional model of Kex1Δp contains all 482 residues; for the purposes of this paper, the first residue will be referred to as Leu1. The $2F_o-F_c$ electron density map is broken at several surface loops: residues 38−45, 58−65, 195−199, 286−290, and 472−477; side chains for many of these residues were not included owing to the poor electron density. The model contains 257 water molecules, with a mean B-factor of 34.0 $Å^2$, and three *N*-acetylglucosamine units, two attached to residue 437 and one to residue 445, with average *B*-factors

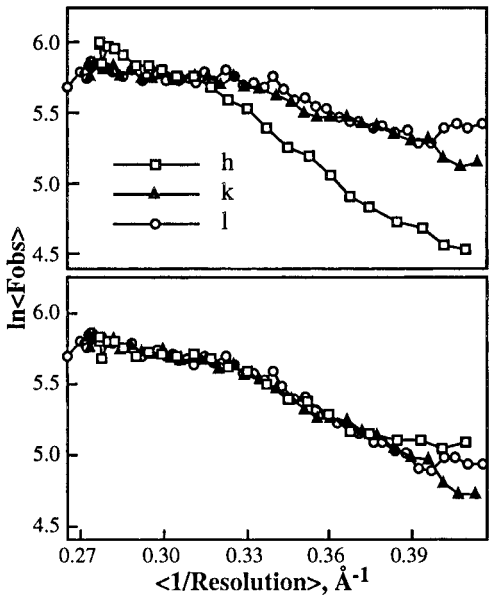

FIGURE 1: Anisotropic diffraction from Kex1Δp crystal. The program DATAMAN (Kleywegt & Jones, 1996) was used for the analysis. Each point on the graph represents a set of reflections with the absolute value of one of the indices held constant; that is, an open square indicates a set of reflections with constant $|h|$; a closed triangle, constant $|k|$; and an open circle, constant $|l|$. For each point (*i.e.*, a set of relections with the absolute value of one index held constant), the average values of 1/resolution and amplitude were calculated and the position of the point on the plot was determined by these two values. The upper graph shows the native data collected at the temperature of 120 K (see Table 1 for details); the lower graph shows the same data after overall anisotropic *B*-factor refinement in X-plor (Brünger, 1992b).

of 51.6 Å^2. After anisotropic refinement of the overall *B*-factor and bulk solvent correction, the model of Kex1Δp has an *R*-factor of 19.7% [19.4% for $F > 2\sigma(F)$] and a free *R*-factor of 25.0% [24.7% for $F > 2\sigma(F)$]. The root mean square deviation from ideal values is 0.006 Å for bond lengths, 1.28° for bond angles, 25.0° for dihedrals, and 1.07° for impropers.

Fold of Kex1Δp. The topology and three-dimensional structure of Kex1Δp are presented in Figures 2 and 3, respectively. The most notable feature of Kex1Δp is its 12-stranded β-sheet: strands $\beta1-\beta8$ belong to the α/β-hydrolase fold; two strands, $\beta9$ and $\beta10$, are inserted between αE and $\beta8$ of the hydrolase fold and are common among CPD-Y, CPD-WII, and HPP; and the last two strands, $\beta11$ and $\beta12$, are formed by residues 469−483 and are unique to Kex1Δp. Strands $\beta1-\beta10$ are surrounded by helices and extended polypeptide chain, forming the core of Kex1Δp that is conserved in CPD-Y and CPDW-II (Figure 3). Residues 211−361 are inserted between $\beta6$ and αD of the α/β-hydrolase fold; this region of the polypeptide sits on top of the Kex1Δp core and will be referred to as the cap. The cap features a helix bundle that is found in the other carboxypeptidases, but the structural conservation in this region is less than that in the core (Figures 2 and 4).

The structure-based sequence alignment (Figure 2b) of Kex1Δp with CPD-Y and CPDW-II can be used to illustrate further the relationship among these three enzymes, and their place in the α/β-hydrolase family. For structural comparison, active-site residues of the three enzymes were superimposed with the program O (Jones *et al.*, 1990) and the superposition was optimized for all CA atoms using the lsq_improve option with a 3.5 Å cutoff. In this way, 319 CA atoms of CPD-Y and 295 atoms of CPDW-II could be matched to corresponding Kex1Δp CA atoms, with rms differences for these atoms of 1.5 and 1.4 Å, respectively; unless otherwise specified, this global superposition will be the basis for all comparisons. The secondary structure elements contained in the α/β-hydrolase cores of the three carboxypeptidases are very well conserved, and deviations of more than 3.5 Å between CA positions are limited to surface loops (Figures 2b and 5). Residues belonging to the catalytic triad—Ser176, His448, and Asp383, as well as the residues contributing to the oxyanion hole, Gly76 and Tyr177—are found within the α/β-hydrolase core, on structurally well-conserved loops (Figure 2).

The topology of the cap is roughly the same for all three carboxypeptidases. As mentioned, the cap is formed primarily by an insertion of residues 211−356 into the α/β-hydrolase fold between $\beta6$ and αD (Figure 2a). In addition, there are two minor elements contributing to the cap: the first is a β-hairpin loop (residues 85−105) that is inserted into αA of the canonical hydrolase fold (Figure 2a). This loop is structurally conserved in CPD-Y and CPDW-II (Figure 2b) but is not present in other α/β-hydrolase enzymes. The second minor cap element is formed from a large loop, residues 126−140, that is present only in Kex1Δp (Figure 2). Those portions of the cap that interact directly with the core, namely, residues 211−230 and 314−353, are as well conserved as the core itself (Figures 2b and 4). The cap contains a helix bundle that is present in all three carboxypeptidases; stabilizing the helix bundle is a disulfide bond (Cys251−Cys271 in Kex1Δp) between two of the helices that is found in the same relative position in Kex1Δp, CPDW-II, and CPD-Y. Despite this topological conservation, the helices are oriented somewhat differently in all three enzymes, and the CA positions of residues within the helix bundle are therefore not well-conserved (Figures 2b and 4).

Cysteine Residues. There are seven cysteine residues in Kex1Δp, six of which form disulfide bridges. Cystine 79−345 forms part of the binding site; it is found in both CPD-Y and CPDW-II and probably helps to stabilize a well-conserved loop that contains one of the oxyanion hole hydrogen-bond donors (NH of Gly76). The rms difference between atoms in cystine 79−345 of Kex1Δp and those in cystine 56−298 of CPD-Y is 1.6 Å; for cystine 56−303 of CPDW-II the value is 1.0 Å. The disulfide bridge connecting two of the helices in the cap domain (cystine 251−271) has an rms difference of 1.5 Å when compared to cystine 217−240 of CPD-Y and 3.1 Å when compared to cystine 210−222 of CPDW-II. The third disulfide bridge is found between residues 293 and 308 in Kex1Δp and may act to stabilize the main chain surrounding S1′; although not truly conserved (there is little sequence conservation in this region of the enzyme), a disulfide bridge is found in the same area for each of CPD-Y and CPDW-II.

Residue 386 is the only free cysteine in Kex1Δp. As expected from the sequence alignment with CPD-Y, this residue was the site of reaction with $HgCl_2$; consistent with the good substitution at this position, $2F_o-F_c$ electron density maps show no evidence of oxidation. Cysteine 386 is located close to the active site histidine (the distance between 386-S$^{\gamma}$ and 448-C$^{\epsilon1}$ is 3.7 Å), on the P1 side of the catalytic triad. This residue is found as Cys341 in CPD-Y, with an rms difference (all atoms) between it and Cys386

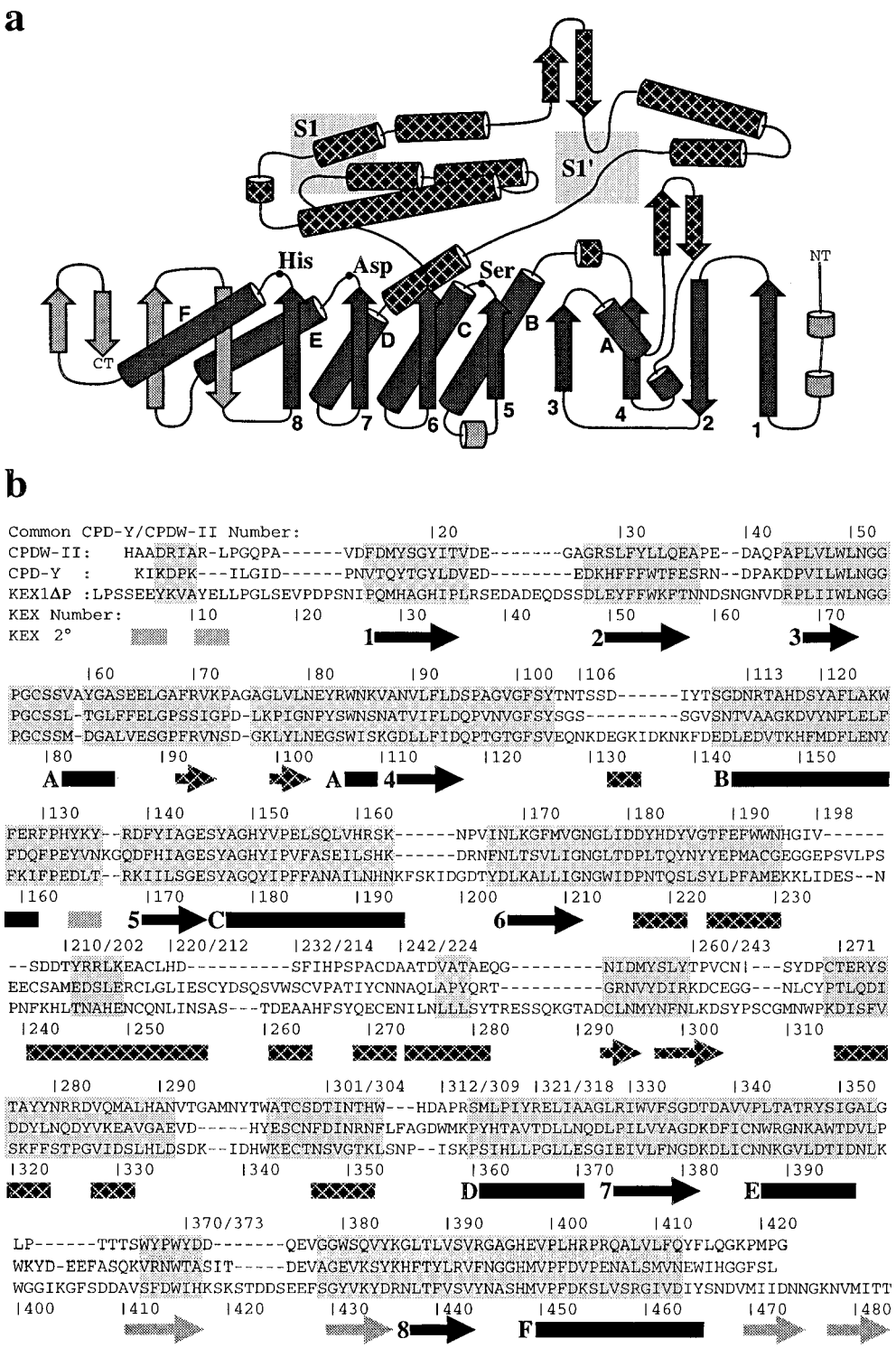

FIGURE 2: (a) Topology of Kex1Δp. Helices are drawn as tubes, while β-strands are drawn as arrows. Secondary structure elements that are common to the α/β-hydrolase fold (Ollis *et al.*, 1992) are heavily shaded (with the α-helices named A−F and β-strands named 1−8), while elements of the cap are filled with a crosshatched pattern. Regions of the cap domain implicated in P1 and P1′ binding are indicated by shading. (b) Structure-based sequence alignment of Kex1Δp with CPDW-II and CPD-Y. The structures were aligned as described in Materials and Methods. The secondary structure of Kex1Δp is indicated and shaded using the same scheme as for panel a. Regions of the sequence that are shaded indicate CA positions that differ by less than 3.5 Å between Kex1Δp and either CPDW-II or CPD-Y. The residue numbering underneath the sequence alignment corresponds to Kex1Δp, while that above corresponds to the common numbering scheme adopted for CPDW-II and CPD-Y (Endrizzi *et al.*, 1994; Liao *et al.*, 1992), except where two numbers are indicated; then the first is for CPD-Y and the second is for CPDW-II.

of Kex1Δp of 0.35 Å. The equivalent residue in CPDW-II is Val341.

Cis-Proline Residues. Prolines at positions 77, 119, 306, and 313 are all in the *cis* conformation. Two of these residues, Pro77 and Pro119, are conserved among all the carboxypeptidases investigated to date (Olesen & Breddam, 1995). Pro77 is part of a functionally important loop (residues 74−81); the NH group of Gly76 is one of the oxyanion hole hydrogen-bond donors, and the carbonyl oxygen of Gly76 is in a good position to accept a hydrogen bond from the main chain of a bound substrate. This loop is stabilized by the structurally conserved disulfide bond

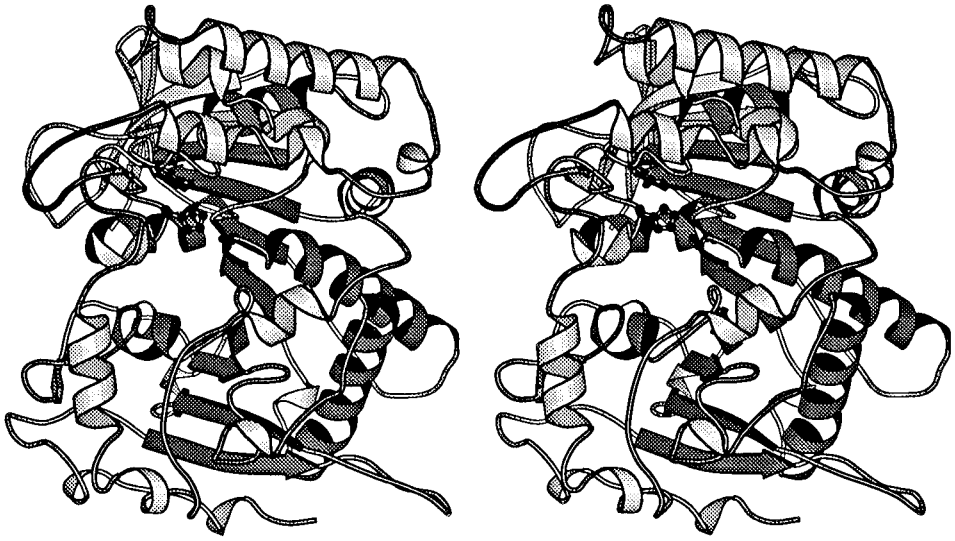

FIGURE 3: Stereo view of the fold of Kex1Δp. Active-site residues—Ser 176, His448, and Asp383—are included as ball-and-stick models. Secondary structure elements common to the α/β-hydrolase family (Ollis *et al.*, 1992) are heavily shaded. The view is from the top, looking onto the cap and into the active site. Figure was produced with the program Molscript (Kraulis, 1991).

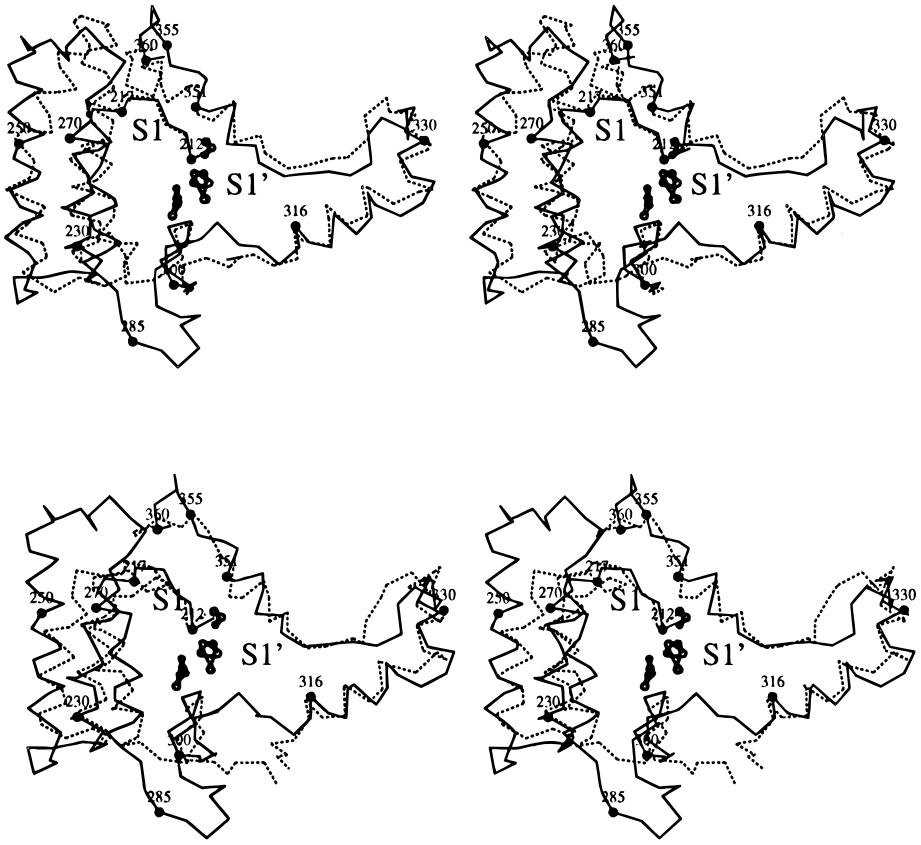

FIGURE 4: Comparison of the Kex1Δp cap to that of CPD-Y or CPDW-II. The active-site residues—Ser176, His448, and Asp383—are included as ball-and-stick models. Subsites S1 and S1′ are indicated. (a, top) Comparison of the Kex1Δp and CPD-Y caps. (b, bottom) Comparison of Kex1Δp and CPDW-II caps. Figures were produced with the program Molscript (Kraulis, 1991).

between Cys79 and Cys345; the presence of the proline residue will add even greater stability to this region. Pro77 is found in the *cis* conformation in both CPD-Y and CPDW-II but not in HPP (Rudenko *et al.*, 1995). Pro119 is present in the *cis* conformation in HPP as well as in CPD-Y and CPDW-II; this residue occurs at the end of β4 in a stretch of polypeptide that packs against the loop mentioned above. The other two *cis*-prolines, Pro306 and Pro313, are not conserved but are present on a sequence of Kex1Δp that is probably an important determinant of its P1′ specificity (see below).

Buried Acid Pairs. There are two pairs of acidic residues that have little or no solvent accessibility, are in largely hydrophobic environments, and whose carboxyl groups are close to each other. The first pair, Glu88 and Glu175, is conserved in all other carboxypeptidases (Olesen & Breddam, 1995) and forms an important part of the catalytic machinery (Stennicke *et al.*, 1996). The distance between Glu88-$O^{\epsilon 1}$ and Glu175-$O^{\epsilon 2}$ is 2.6 Å, and between Glu88-$O^{\epsilon 2}$ and Glu175-$O^{\epsilon 1}$ it is 3.6 Å, indicating that there are probably hydrogen bonds between the carboxyl-group oxygens, and, since the crystals were grown at pH 6.5, either one or both

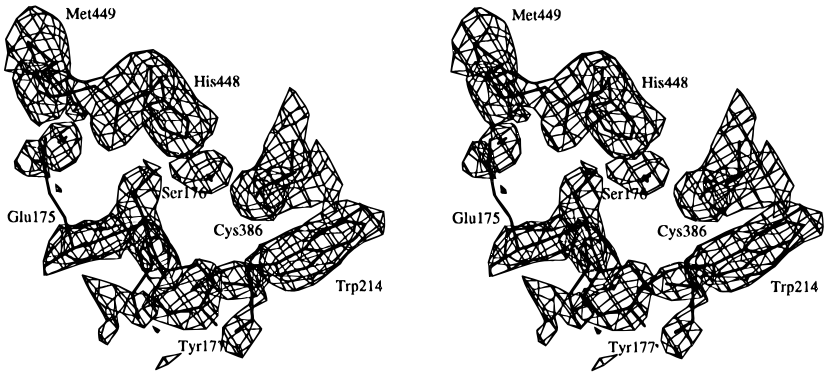

FIGURE 5: Active site of Kex1Δp with an electron density omit map. The final model of Kex1Δp was subjected to simulated annealing, with Ser176 and all residues within a 4 Å radius omitted from the structure factor calculations (Hodel *et al.*, 1992). The electron density map was produced using coefficients F_o-F_c and α_{calc}; the calculated structure factors and phases were derived from the annealed structure. Atoms from the final refined model of Kex1Δp are included.

of these residues may have a perturbed pK_a. While there was very clear electron density for the side chains of both of these residues, it is noteworthy that in the structure of CPDW-II with a bound aldehyde and free arginine in S1′ (Bullock *et al.*, 1996) the side chain corresponding to Glu175 is rotated 120° about χ1 (Figure 6). There is space to accommodate such a conformation in Kex1Δp, but in the current model this space is occupied by a water molecule. Such a change in the side-chain conformation of Glu175 does not substantially increase the separation between its carboxyl group and that of Glu88 but attempts to model a substrate into the active site of Kex1Δp indicate that this alternative conformation of Glu175 may be required for binding of the substrate carboxyl group.

The other pair of buried acidic residues consists of Asp83 and Glu344. The carboxyl group of Asp83 is solvent-inaccessible and is probably protonated: both $O^{\delta1}$ and $O^{\delta2}$ of Asp83 interact with $O^{\epsilon1}$ of Glu344, which is also solvent-inaccessible. Glu344-$O^{\epsilon2}$ is exposed to solvent and well-positioned to provide a negative charge to compensate for the positively charged basic group of a P1′ side chain.

The Active Site. Throughout model-building and refinement there was very clear electron density for all of the active-site residues; however, there was also electron density that could not be explained by the model of Kex1Δp alone. This electron density was apparent in maps produced with MIR phases as well as maps calculated with combined phases from MIR and the molecular replacement solution. The Kex1Δp model was fully refined without addition of water molecules to this region, but the electron density persisted. However, when simulated annealing omit maps were calculated (omitting Ser176 and residues within a 4 Å radius; Hodel *et al.*, 1992), this electron density was no longer present (Figure 5); these electron density omit maps, calculated with either $2F_o-F_c$ or F_o-F_c coefficients, clearly indicated that the Ser176 side chain was hydrogen-bonded to the active-site histidine, His448 (Figure 5). It is likely that the active-site serine of Kex1Δp forms a covalent bond (possibly with acetate ions, which are present in the crystallization buffer at a concentration of 200 mM) in some fraction of the molecules, while in the rest of the molecules this area is occupied by water molecules that are not covalently bonded to the serine. In the final Kex1Δp model, O^{γ} of Ser176 is hydrogen-bonded to $N^{\epsilon2}$ of His448, and the region immediately surrounding O^{γ} contains three water molecules.

As with CPD-Y and CPDW-II, only one of the carboxylate oxygens ($O^{\delta2}$) of Asp383 is coplanar with the imidazole ring of His448, with a distance between $O^{\delta2}$ and $N^{\delta1}$ of 2.9 Å. The side chains of the active-site residues (Ser176, Asp383, and His448) of Kex1Δp can be superimposed (locally) over those from CPD-Y and CPDW-II with rms deviations of 0.455 and 0.484 Å, respectively. If the atoms that determine the hydrogen-bonding characteristics of the oxyanion hole—namely, C, O, and N forming the peptide bonds between residues 75−76 and 176−177—are included, then the rms deviations become 0.442 Å for CPD-Y and 0.480 Å for CPDW-II. Thus, the catalytic machinery is the same for all three enzymes, and the differences that determine their substrate preferences must be found in the subsites responsible for binding the substrate side chains.

Subsite 1′. S1′ of Kex1Δp includes residues 304−314, a region for which there are significant main chain differences between the three enzymes. These residues form a wall at one end of the S1′ site; this wall is present in CPD-Y, but in CPDW-II the main chain in this region is cleaved during proenzyme activation (Breddam *et al.*, 1987), and although there is a disulfide bond connecting the two chains, the area is relatively open (Figure 6). The proline in the Trp312-Pro313 sequence of Kex1Δp adopts a *cis* conformation, with the aliphatic ring stacking against the indole ring of Trp312. These residues fill a space that is empty in CPDW-II and is occupied by Y269 in CPD-Y (Figure 6). The sequence from 304 to 311 contains three extra residues when compared to the equivalent sequence from CPD-Y (Figure 2b), and it can be seen (Figure 6) that these residues in Kex1Δp occupy an area that is empty in the CPD-Y S1′ site. Thus, the S1′ site in Kex1Δp is more confined than the equivalent sites in either CPDW-II or CPD-Y.

In terms of an overall structure, the S1′ of Kex1Δp more closely resembles that of CPD-Y than that of CPDW-II. On the other hand, charged residues are absent from S1′ of CPD-Y, whereas both Kex1Δp and CPDW-II contain S1′ acidic groups that are presumably important for binding of P1′ basic residues. In Kex1Δp, Asp83 and Glu344 take the place of neutral residues found in both CPD-Y (Thr60 and Ser297) and CPDW-II (Tyr60 and Thr302). The three glutamic acid residues that are probably important for binding of basic side chains in CPDW-II, namely Glu64, Glu272, and Glu398, are present as hydrophobic residues in both Kex1Δp (Val87, Ile316, and Met449) and CPD-Y (Phe64, Leu272, and Met398).

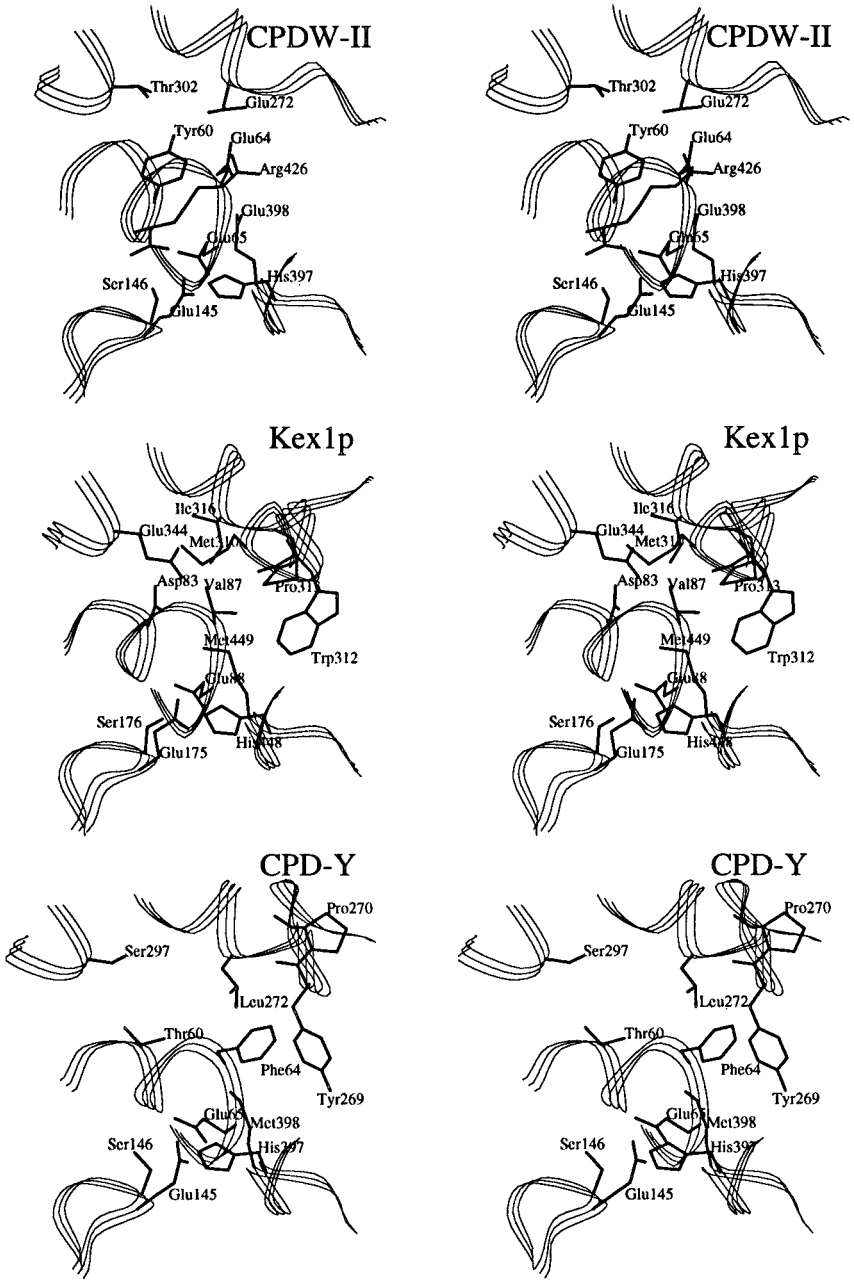

FIGURE 6: Comparison of the P1′ binding sites of the serine carboxypeptidases. S1′ of Kex1Δp is the middle diagram; at the top is the structure of CPDW-II, with bound arginine, and at the bottom is S1′ of CPD-Y.

Subsite 1. The base or foundation of S1 is formed by residues from the core: Tyr177, Ile315, *cis*-Pro77, and Trp214 (Figure 7). Resting on this base are elements from the cap, and since the helix bundle in the cap is closer in structure to that of CPD-Y (Figure 4), S1 of Kex1Δp bears a closer resemblance to CPD-Y than to CPDW-II. In particular, the helix composed of residues Gln269−Tyr281 has a similar length and orientation as the corresponding helix in CPD-Y (residues Ile238−Arg250; Figure 7). When Kex1Δp and CPD-Y are compared, main-chain differences affecting S1 involve two sections—one containing residues Ala264−Glu270 and the other containing residues Lys353−Pro360 (Figure 7); the first section connects two of the cap helices, making an excursion into S1, and it is partially covered by the second section, which tends to point outward into the solvent. The situation is reversed in CPD-Y, with residues 304−315 (corresponding to 353−360 of Kex1Δp) forming part of S1 (Figure 7). These main-chain changes mean that residue Trp312 of CPD-Y, which increases the P1 Lys/Leu preference of CPD-Y when mutated to Asp or Glu (Olesen & Breddam, 1995), is not present in S1 of Kex1Δp (see below).

There are two notable side-chain differences between Kex1Δp and CPD-Y that may contribute to differences in P1 specificity. The first change is the introduction in Kex1p of an acidic side chain, Glu272, that is well-positioned to interact with a basic side chain of the substrate. There is no charged residue in an equivalent position of CPD-Y; however, when Trp312 is mutated to Asp or Glu (a change that increases the P1 Lys/Leu preference of the enzyme; Olesen & Breddam, 1995), the carboxylic acid group would occupy a position roughly equivalent to that of Glu272 of Kex1Δp. The second side-chain substitution occurs at Trp214, which corresponds to Leu178 of CPD-Y. It has been shown (Olesen & Breddam, 1995) that enzymes with

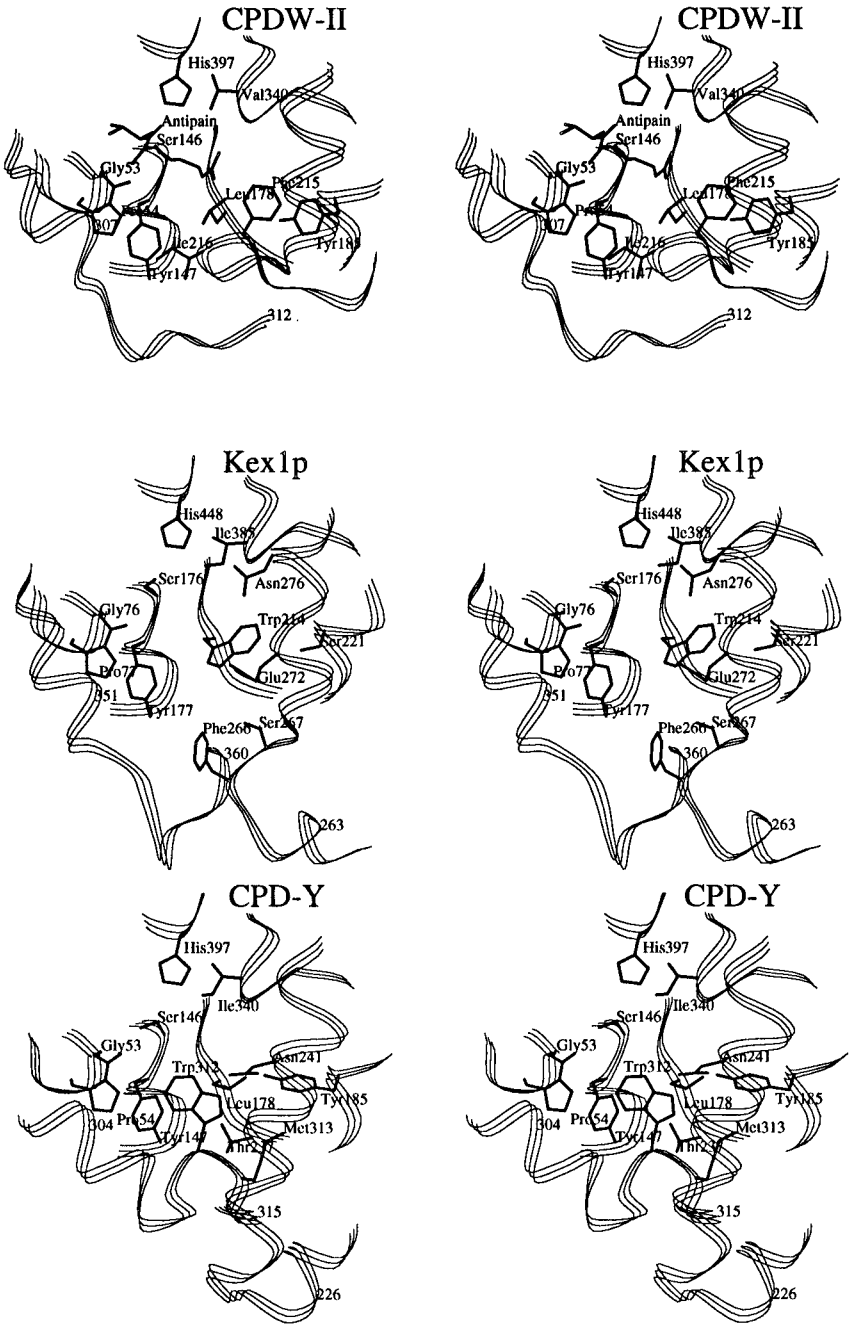

FIGURE 7: Comparison of the P1 binding site of the serine carboxypeptidases. S1 of Kex1Δp is the middle diagram; at the top is the structure of CPDW-II, with an aldehyde, antipain, bound covalently to the active-site serine. P1 of the substrate for Kex1Δp is expected to occupy roughly the same position as the arginine side chain of antipain in the CPDW-II structure. Below Kex1Δp is S1 of CPD-Y.

aromatic amino acids at this position demonstrate a preference for basic residues.

DISCUSSION

Analysis of the primary structures of 30 serine carboxypeptidases led Olesen and Breddam (1995) to divide the enzymes into two groups: they predicted that the enzymes in one group would most closely resemble CPD-Y in their three-dimensional structures, while those in the second group would resemble CPDW-II. Carboxypeptidases in the CPDW-II group are cleaved between residues 260 and 270 (CPD-Y numbering) during maturation in the secretory pathway, and they demonstrate a preference for basic residues at P1 (Olesen & Breddam, 1995). The CPD-Y group consists of proteins with a single (uncleaved) polypeptide chain and no consistent substrate preference at P1. HPP was grouped with CPDW-II, and although the coordinates for the HPP structure (Rudenko *et al.*, 1995) are not yet publicly available, it appears that the enzyme more closely resembles CPDW-II than CPD-Y. The catalytic domain of Kex1p was grouped with CPD-Y, and it is clear from our model of Kex1Δp that this placement is correct.

The core of the serine carboxypeptidases is based upon the same α/β-hydrolase fold found in the lipases, acetylcholinesterase, haloalkane dehalogenase, and others (Ollis *et al.*, 1992) and provides a rigid scaffold that correctly positions the common catalytic elements: the catalytic triad, oxyanion hole, and the glutamic acid residues (Glu88 and Glu175 in Kex1Δp) that are important for the hydrolysis of peptides (Stennicke *et al.*, 1996). On the other hand, many

of the elements determining substrate specificity come from the cap. In the cases of HPP and CPDW-II, the helix bundle in the cap forms part of the dimer interface (Liao *et al.*, 1992; Rudenko *et al.*, 1995), and sequences in the cap are important in maturation. Thus, the caps of the serine carboxypeptidases have allowed different functions to evolve based on the same catalytic core.

The role of Kex1p is to complete the maturation of prohormones to biologically active peptides, and on this basis, one would expect that proteolytic cleavage by Kex1p would not continue past C-terminal basic residues. In the case of three peptides, namely, α-pheromone, [Leu^5]-enkephalin, and [Met^5]-enkephalin, Kex1p did not cleave past C-terminal basic residues, even after prolonged incubation (Latchinian-Sadek & Thomas, 1993). There is *in vivo* evidence that Kex1p does not cleave beyond the C-terminal basic residues of K1 killer toxin (Zhu *et al.*, 1987). The structure of Kex1Δp offers some clues as to its specificity and why it is more specific than either CPDW-II or CPD-Y. The preference of Kex1p for basic residues at P1′ can be explained by the presence of complementary acidic residues in S1′. Thus, both Asp83 and Glu344 of Kex1p will provide a negatively charged environment that would be expected to facilitate binding of a positively charged side chain. The presence of these acidic groups is not enough to explain the high specificity of Kex: S1′ of CPDW-II contains three acidic residues that are probably important for binding P1′ basic side chains, and yet CPDW-II does not exhibit an absolute requirement for basic residues at the C-terminus (Breddam *et al.*, 1987).

Another factor that may contribute to the high specificity of Kex1p is that S1′ is sterically restricted when compared to S1′ of CPDW-II. Among other differences, the CPDW-II polypeptide undergoes cleavage during maturation, and the section of the chain that is cleaved corresponds roughly to the section of Kex1Δp that partially occludes S1′. While S1′ of CPDW-II is open with a high degree of solvent accessibility, inspection of the solvent-accessible surface of Kex1Δp S1′ indicates an almost perfect fit for an arginine or lysine side chain. This shape complementarity aided by the presence of the negative charges can explain why Kex1p will not catalyze peptide-bond cleavage with large hydrophobic or acidic residues at the C-terminus of the substrate.

With alanine at the P1 position, CPDW-II is able to cleave a variety of amino acids from the C-terminus of dipeptide substrates. For example, the k_{cat}/K_m (min^{-1} mM^{-1}; pH 4.5) of CPDW-II for Z-Ala-Arg is 52 000 min^{-1} mM^{-1}, while for Z-Ala-Phe, Z-Ala-Met, and Z-Ala-Ala the values are 4200, 7400, and 1000 min^{-1} mM^{-1}, respectively (Breddam *et al.*, 1987). Given these data for CPDW-II and the nature of S1′ in Kex1Δp, there is no obvious reason why Kex1Δp should not be able to cleave alanine from the C-terminus of a substrate. However, the limited data available indicate that C-terminal trimming of α-pheromone, K1 killer toxin, and the enkephalins (sequences, Table 3; Latchinian-Sadek & Thomas, 1993; Zhu *et al.*, 1987) does not proceed past the basic residues. Thus, factors other than the structure of S1′ must contribute to the high specificity of Kex1Δp, and therefore the P1 preference and the sequences of the substrates should also be considered.

The P1 preference for Kex1p has not yet been investigated in detail, but preliminary data (Klaus Breddam, personnal communication) indicate that its P1 Lys/Leu preference is relatively high. Mutagenesis studies of CPD-Y, which most closely resembles the structure of Kex1Δp in the S1 region, provide an explanation for the high Lys/Leu preference of Kex1p. Wild-type CPD-Y demonstrates a preference for hydrophobic residues at P1; however, it was shown that substitution of tryptophan for leucine at position 178 of CPD-Y, a change that is found in Kex1p, increased the Lys/Leu preference of CPD-Y 50-fold (Olesen *et al.*, 1994). More importantly, S1 of Kex1Δp includes an acidic residue, Glu272. In the case of CPD-Y, introduction of an acidic residue into S1 resulted in an 1800−2800-fold increase in the P1 Lys/Leu preference of the enzyme (Olesen *et al.*, 1994). Therefore, the presence of Glu272 and Trp178 in Kex1p are likely factors in its high P1 Lys/Leu preference.

Table 3: Peptide Substrates of Kex1

peptide	sequence	references
α-pheromone	WHWLQLKPGGPMY-**KR**	Kurjan & Herskowitz, 1982; Latchinian-Sadek & Thomas, 1993
K1 toxin	.....PNGATVA-**RR**	Skipper et al., 1984; Zhu et al., 1987
K2 toxin	.....DNSTGIV-**KR**	Dignard et al., 1991
Leucine Enkalphalin	YGGFL-**(K/R)R**	Latchinian-Sadek & Thomas, 1993
Methionine Enkalphalin	YGGFM-**RR**	Latchinian-Sadek & Thomas, 1993

Given that Kex1p shows a great preference for basic residues at P1′ and demonstrates a preference for basic residues at P1, the sequences of the peptide substrates (Table 3) may play an important role in determining the specificity of the enzyme: it could be that the natural substrates of Kex1p are ideally designed to stop cleavage after the removal of the two basic residues. That is, although CPD-Y exhibits a broad specificity, there are some sequences through which it cannot cleave or through which cleavage is extremely slow (Hayashi, 1977). The mature peptides in Table 3 all end in a sequence of hydrophobic residues, and given the negative charges in both S1 and S1′ of Kex1Δp, these substrates may be particularly unsuitable for cleavage by Kex1p. Kinetic studies are underway to characterize further Kex1Δp and to address these possibilities.

REFERENCES

Breddam, K., Sørensen, S. B., & Svendsen, I. B. (1987) *Carlsberg Res. Commun. 52*, 297−311.

Brünger, A. T. (1992a) *Nature 355*, 472−475.

Brünger, A. T. (1992b) *XPLOR: A System for X-ray Crystallography and NMR, Version 3.1*, Yale University Press, New Haven, CT.

Brünger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458−460.

Bryant, N. J., & Boyd, A. (1993) *J. Cell Sci. 106*, 815−822.

Bullock, T. L., Breddam, K., & Remington, S. J. (1996) *J. Mol. Biol. 255*, 714−725.

CCP4 (1994) *Acta Crystallogr. D50*, 760−763.

Cooper, A., & Bussey, H. (1989) *Mol. Cell. Biol. 9*, 2706−2714.

Cygler, M., Schrag, J. D., Sussman, J. L., Harel, M., Silman, I., Gentry, M. K., & Doctor, B. P. (1993) *Protein Sci. 2*, 366−382.

Denault, J.-B., & Leduc, R. (1996) *FEBS Lett. 379*, 113−116.

Dignard, D., Whiteway, M., Germain, D., Tessier, D., & Thomas, D. Y. (1991) *Mol. Gen. Genet. 227*, 127−136.

Dmochowska, A., Dignard, D., Henning, D., Thomas, D. Y., & Bussey, H. (1987) *Cell 50*, 573−584.

Endrizzi, J. A., Breddam, K., & Remington, S. J. (1994) *Biochemistry 33*, 11106−11120.

Fuller, R. S., Brake, A. J., & Thorner, J. (1989) *Science 246*, 482−486.
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.
Galjart, N. J., Morreau, H., Willemsen, R., Gilemans, N., Bonten, E. J., & d'Azzo, A. (1991) *J. Biol. Chem. 266*, 14754−14762.
Germain, D., Zollinger, L., Racine, C., Gossard, F., Dignard, D., Thomas, D. Y., Crine, P., & Boileau, G. (1990) *Mol. Endocrinol. 4*, 1572−1579.
Hayashi, R. (1977) *Methods Enzymol. 47*, 84−93.
Hayashi, R., Bai, Y., & Hata, T. (1975) *J. Biochem. (Tokyo) 77*, 69−79.
Hodel, A., Kim, S.-H. A. T., & Brunger, A. T. (1992) *Acta Crystallogr. A48*, 851−859.
Hutton, J. C. (1990) *Curr. Opin. Cell Biol. 2*, 1131−1142.
Jackman, H. L., Tan, F., Tamei, H., Beurling-Harbury, C., Li, X.-L., Skidgel, R. A., & Erdös, E. G. (1990) *J. Biol. Chem. 265*, 11265−11272.
Jones, T. A., Bergdoll, M., & Kjeldgaard, M. (1990) *O: a macromolecular modeling environment*, Springer-Verlag, New York.
Kleywegt, G. J., & Jones, T. A. (1996) *Acta Crystallogr. D52*, 826−828.
Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Kurjan, J., & Herskowitz, I. (1982) *Cell 30*, 933−943.
Latchinian-Sadek, L., & Thomas, D. Y. (1993) *J. Biol. Chem. 268*, 534−540.
Latchinian-Sadek, L., & Thomas, D. Y. (1994) *Eur. J. Biochem. 219*, 647−652.
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.
Mortensen, U. H., Remington, S. J., & Breddam, K. (1994) *Biochemistry 33*, 508−517.
Navaza, J. (1994) *Acta Crystallogr. A50*, 157−163.
Olesen, K., & Breddam, K. (1995) *Biochemistry 34*, 15689−15699.
Olesen, K., Mortensen, U. H., Aasmul-Olsen, S., Kielland-Brandt, M. C., Remington, S. J., & Breddam, K. (1994) *Biochemistry 33*, 11121−11126.
Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J. L., Verschueren, K. H. G., & Goldman, A. (1992) *Protein Eng. 5*, 197−211.
Otwinowski, Z. (1991) in *Proceedings of the CCP4 Study Weekend*, 29−30 January 1991, Data Collection and Processing (Sawyer, L., Isaacs, N., & Burley, S., Eds.) pp 56−62, SERC Daresbury Laboratory, Daresbury, Warrington, U.K.
Pshezhetsky, A. V., Vinogradova, M. V., Elsliger, M.-A., El-Zein, F., Svedas, V. K., & Potier, M. (1995) *Anal. Biochem. 230*, 303−307.
Rudenko, G., Bonten, E., d'Azzo, A., & Hol, W. G. J. (1995) *Structure 3*, 1249−1259.
Seidah, N. G., Chretien, M., & Day, R. (1994) *Biochimie 76*, 197−209.
Shilton, B. H., Li, Y., Tessier, D., Thomas, D. Y., & Cygler, M. (1996) *Protein Sci. 5*, 395−397.
Skidgel, R. A. (1988) *Trends Pharmacol. Science 9*, 299−304.
Skipper, N., Thomas, D. Y., & Lau, P. C. K. (1984) *EMBO J. 3*, 107−111.
Stennicke, H. R., Mortensen, U. H., & Breddam, K. (1996) *Biochemistry 35*, 7131−7141.
Tan, F., Morris, P. W., Skidgel, R. A., & Erdš, E. G. (1993) *J. Biol. Chem. 268*, 16631−16638.
Thomas, L., Cooper, A., Bussey, H., & Thomas, G. (1990) *J. Biol. Chem. 265*, 10821−10824.
Wickner, R. B., & Leibowitz, M. J. (1976) *Genetics 82*, 429−442.
Wishart, D. S., Boyko, R. F., Willard, L., Richards, F. M., & Sykes, B. D. (1994) *Comput. Appl. Biosci. 10*, 121−132.
Zhu, H., Bussey, H., Thomas, D. Y., Gagnon, J., & Bell, A. W. (1987) *J. Biol. Chem. 262*, 10728−10732.
Zollinger, L., Racine, C., Crine, P., Boileau, G., Germain, D., & Thomas, D. Y. (1990) *Biochem. Cell. Biol. 68*, 635−640.

BI970433N