

2.4 Å Resolution Crystal Structure of the Prototypical Hormone-Processing Protease Kex2 in Complex with an Ala-Lys-Arg Boronic Acid Inhibitor^{†,‡}

Todd Holyoak,[§] Mark A. Wilson,[§] Timothy D. Fenn,[§] Charles A. Kettner,^{||} Gregory A. Petsko,[§]
Robert S. Fuller,[⊥] and Dagmar Ringe^{*,§}

Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02454,
Dupont Pharmaceuticals Company, Wilmington, Delaware, 19880, and Department of Biological Chemistry,
University of Michigan Medical School, Ann Arbor, Michigan, 48109

Received March 18, 2003; Revised Manuscript Received April 10, 2003

ABSTRACT: This paper reports the first structure of a member of the Kex2/furin family of eukaryotic pro-protein processing proteases, which cleave sites consisting of pairs or clusters of basic residues. Reported is the 2.4 Å resolution crystal structure of the two-domain protein ssKex2 in complex with an Ac-Ala-Lys-boroArg inhibitor ($R = 20.9\%$, $R_{\text{free}} = 24.5\%$). The Kex2 proteolytic domain is similar in its global fold to the subtilisin-like superfamily of degradative proteases. Analysis of the complex provides a structural basis for the extreme selectivity of this enzyme family that has evolved from a nonspecific subtilisin-like ancestor. The P-domain of ssKex2 has a novel jelly roll like fold consisting of nine β strands and may potentially be involved, along with the buried Ca^{2+} ion, in creating the highly determined binding site for P₁ arginine.

Kex2 (kexin; E. C. 3.4.21.61) is a Ca^{2+} -dependent trans-membrane protease found in the yeast *Saccharomyces cerevisiae* whose name derives from the killer expression phenotype of Kex2 mutant cells (1–3). In *S. cerevisiae*, Kex2 is necessary for production and secretion of mature α -factor and killer toxin by proteolysis at paired dibasic sites (4). Pleiotropic effects of deleting the *KEX2* gene suggest the existence of additional targets (5, 6) that likely include cell wall proteins and enzymes (7, 8). The Kex2 homologues in pathogenic fungi *Candida albicans* and *Candida glabrata* are virulence factors (9, 10) with the *albicans* molecule implicated in the processing of at least 33 additional proteins (9). Kex2 is the prototype of a large family of eukaryotic pro-protein processing proteases that includes furin, PC2, PC3/PC1, PC4, PACE4, PC5/6, and PC7/LPC in mammals (11). This family of proteases is responsible for the processing of virtually all neuropeptides and peptide hormones as well as proinsulin, coagulation factors, and many growth factors and their receptors (12). These serine proteases comprise a discrete branch of the subtilisin superfamily spanning eukaryotes from yeast to humans (13). What distinguishes the Kex2 family of pro-protein convertases from subtilisin and its homologues is a high specificity for cleavage C-terminal to paired basic sites, most often KR or RR.

Previous biochemical investigations have characterized the specificity and enzymatic mechanism of Kex2 and furin (see ref 14 for a comprehensive review). These enzymes follow the classic serine protease mechanism, possessing the archetypal catalytic triad of serine, histidine, and aspartate residues. Consistent with the classical mechanism, Kex2 has been demonstrated to utilize a conserved asparagine residue located in the putative oxyanion hole, thereby stabilizing the negative charge that develops on the scissile carbonyl bond during the transition state (13, 15, 16).

Although considerable insight has been gained into the specificity of Kex2 and furin, little is known about the structural basis for the selectivity of these enzymes. Biochemical characterization of Kex2 suggests that the P₁ position is the primary specificity determinant while P₂ and P₄ are important energetically but to a lesser degree (14, 17). (Substrate residues are designated in accordance with the naming convention of Schechter and Berger, with the hydrolyzed bond lying between P₁ and P_{1'} (18).) Furin, however, seems to generate most of its selectivity through interactions with both P₁ and P₄ (14, 19, 20). This difference between Kex2 and furin specificity determinants seems to represent two discrete subsets within the eukaryotic family of pro-protein processing proteases. Residues C-terminal to the cleavage site seem to be unimportant for substrate recognition; however, it has been suggested that Kex2 may disfavor large residues at P_{1'} (21). Kex2 and the family of protein convertases also show an absolute dependence on Ca^{2+} for activity (3). Subtilisin and other degradative subtilases (i.e., thermolysin and proteinase K) also bind Ca^{2+} ; however, Ca^{2+} is not necessary for subtilisin activity and has been suggested to act purely as a stabilizer of tertiary structure, thereby enhancing thermostability (22). The role of calcium in the mechanism of Kex2 and the related family members is therefore unknown.

[†] This work was supported in part by a N.I.H. Grant GM39697 (to R.S.F.) and by a supplement to that grant (to D.R.) and in part by a grant from the Lucille P. Markey Charitable Trust to Brandeis University.

^{*} Corresponding author. E-mail: ringe@brandeis.edu.

[‡] Coordinates and structure factors have been deposited in the RCSB protein databank (<http://www.rcsb.org>) under the accession code 1OT5.

[§] Brandeis University.

^{||} Retired from Dupont Pharmaceuticals Company.

[⊥] University of Michigan Medical School.

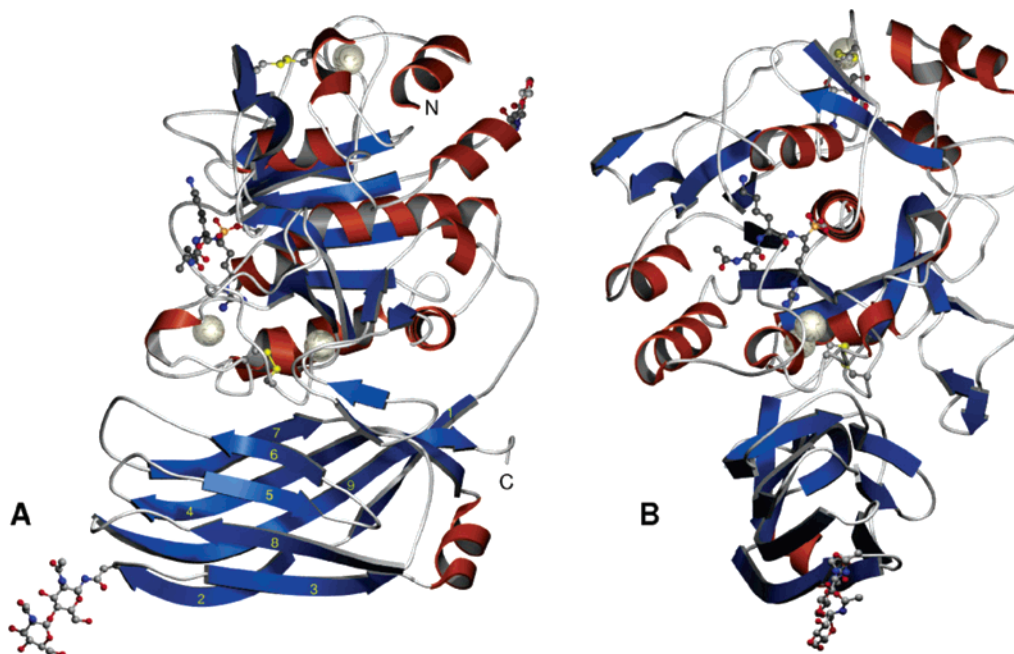


FIGURE 1: Ribbon diagram representation of the monomer of ssKex2. (A) The Glc-Nac residues, disulfide bonds, catalytic triad, and inhibitor are shown as ball-and-stick representations and the strands comprising the P-domain are numbered. (B) A rotation about the Y axis of 90° from panel A. All figures were generated using POVscript+ (<http://www.brandeis.edu/~fenn/povscript>) (56) and rendered using POVray (<http://www.povray.org>).

Kex2 is a multidomain protein containing an N-terminal signal peptide that is responsible for directing the protein into the endoplasmic reticulum, with its ultimate destination being the trans-Golgi network (23). Immediately following the signal peptide is the pro-domain. This domain has been shown in subtilisins to be important in protein folding, and its removal is necessary for enzyme activation (24, 25). In Kex2, the pro-domain is cleaved from the remainder of the enzyme through an intramolecular self-mediated cleavage event carboxyl to R109 (26, 27). This is followed by further processing in the Golgi by Ste13 dipeptidyl aminopeptidase that removes two N-terminal dipeptides (L110-P111 and V112-P113) generating the mature N-terminus beginning at A114 (28). The subtilisin-like domain follows the pro-domain and has approximately 30% identity with subtilisin BPN'. Following the subtilisin domain is the ubiquitous P (Homo B)-domain, an element in all members of the pro-protein convertase family of enzymes characterized to date but in none of the degradative subtilisins. The function of this domain is unknown; however, it has been suggested to be necessary for catalytic function in vivo (29). In Kex2, the P-domain is followed by a serine/threonine rich region that is hyper O-glycosylated. A single trans-membrane spanning helix and cytosolic tail complete the enzyme architecture. In mammals and lower animals, there appear to be two subclasses of the family, those with a transmembrane domain (furin) and those lacking it (PC1/3, PC2) (14).

The enzyme form utilized in prior biochemical studies and under study in this report, ssKex2¹ is a fully active construct that lacks the serine/threonine rich region and the transmem-

brane domain (28, 30). It is expressed in the yeast strain ASY1 that lacks the *STE13* gene to reduce any potential microheterogeneity at the N-terminus. This construct is expressed as a soluble protein that is secreted from the expressing yeast cell generating a molecule comprising 507 amino acids and N-linked glycans at two of three potential N-linked glycosylation sites (28).

This paper reports the 2.4 Å resolution crystal structure of ssKex2 in complex with a peptidyl boronic acid inhibitor, the first structure of a pro-protein processing protease of the Kex2/furin family. This structure gives insight into the structural determinants of the high sequence specificity of Kex2, which contrasts with the promiscuous nature of the subtilisin-like proteases from which this family has evolved. A catalytic role for the essential calcium ion is determined, as well as a novel jelly roll like fold exhibited by the P-domain of this enzyme.

MATERIALS AND METHODS

Materials. Bis-Tris buffer was purchased from Research Organics (Cleveland, OH). The substrate Boc-Gln-Arg-Arg-AMC was purchased from Bachem (Bachem A., Switzerland). The boronic acid inhibitor, Ac-Ala-Lys-boroArg-pinanediol (**1**, Figure 2B) was a generous gift of Dr. Charles Kettner of DuPont Pharmaceuticals, Wilmington, DE whose synthesis has been described elsewhere (31).

Crystallographic materials and screening kits were purchased from Hampton Research (Laguna Niguel, CA). Malonic acid was purchased from Sigma (St. Louis, MO), and a saturated solution at pH 7.2 was prepared as previously described (32). All other reagents were of the highest purity available.

Kex2 Expression and Purification. Kex2 protein concentration was determined using a calculated extinction coefficient of $\epsilon_{280} = 0.595 \text{ mL/mg}$ (Swiss-Prot, <http://us.expasy.org/>

¹ Abbreviations: ASU, asymmetric unit; DTT, dithiothreitol; Glc-Nac, *N*-acetyl-glucosamine; Hg-PNP, 2-chloromercuri-4-nitrophenol; MAD, multiple wavelength anomalous diffraction; NCS, noncrystallographic symmetry; SAD, single wavelength anomalous diffraction; SIRAS, single isomorphous replacement with anomalous scattering; ssKex2, secreted soluble Kex2.

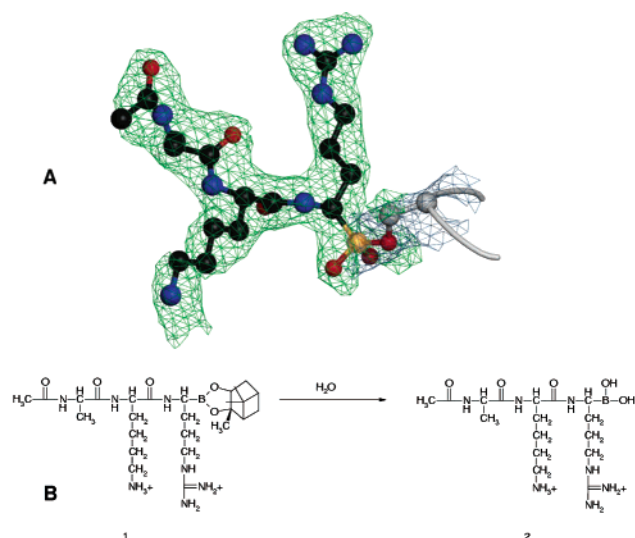


FIGURE 2: Ac-Ala-Lys_boroArg inhibitor of ssKex2. (A) The $2F_o - F_c$ and $F_o - F_c$ electron density maps for the bound boronic acid inhibitor and S385 prior to inclusion of the inhibitor into the model. Shown in blue is the $2F_o - F_c$ map rendered at 1σ and in green the $F_o - F_c$ map at 2σ . The carbon atoms of the inhibitor are shown in black, while the enzyme carbon atoms are rendered in light gray. (B) Line drawing representation of Ac-Ala-Lys_boroArg-pinane diol inhibitor **1**, which upon addition to aqueous solvent becomes the active boronic acid species **2**.

sprot/). This value was found to compare favorably with that determined using the BioRad and Pierce colorimetric methods.

ss-Kex2 was prepared as previously described (30) with the following exceptions. The fermentation temperature was 25 and not 30 °C, and growth times were extended approximately 1 day to accommodate the lower temperature. Cell medium was diluted 4-fold with cold 40 mM Bis-Tris base and stirred for 1 h with Fast-Flow Q-Sepharose (Pharmacia, Piscataway, NJ). The resin was harvested by filtration through a large sintered glass funnel, via vacuum suction. The resin bed was subsequently washed with 5–10 column volumes of 40 mM Bis-Tris pH 7.2, 10 mM NaCl, 2 mM CaCl₂ buffer and loaded into a glass column (5 × 25 cm). The enzyme was eluted with a linear gradient (5 column volumes) of 10–500 mM NaCl in the same buffer. Fractions exhibiting Kex2 activity were determined by a qualitative assay developed previously (33). In brief, a 96-well plate is used into which 100 μ L of crude protein fraction is pipetted along with 200 μ L of reaction mix (200 mM bis-Tris pH 7.2, 1 mM CaCl₂, 0.01% v/v Triton X-100, 0.5% v/v DMSO, 200 μ M Boc-Gln-Arg-Arg-AMC). Those wells that showed immediate fluorescence under a long wavelength UV light were pooled and concentrated to 10 mL in an Amicon nitrogen concentrator utilizing a YM-30 membrane (Millipore Corporation, Bedford, MA). Several rounds of concentration and dilution with 40 mM Bis-Tris, pH 7.2, 10 mM NaCl, 2 mM CaCl₂ were performed to reduce the sodium chloride concentration and remove a colored contaminant.

The final concentrate was subsequently loaded onto a BioScale Q20 column (BioRad, Hercules, CA) equilibrated with 40 mM Bis-Tris pH 7.2, 10 mM NaCl, 2 mM CaCl₂ buffer. After loading, the protein was washed with five column volumes of 40 mM Bis-Tris pH 7.2, 10 mM NaCl, 2 mM CaCl₂ buffer and eluted with a gradient from 10 to 500 mM NaCl.

Those fractions containing Kex2 activity were pooled and concentrated in an Amicon concentrator. Kex2 was subsequently inactivated using a 4-fold molar excess of Ac-Ala-Lys-boroArg inhibitor (Figure 2B) dissolved in DMSO. The inactivation mix was allowed to incubate overnight at 4 °C and was subsequently assayed to ensure no enzyme activity remained. The protein solution was concentrated further to a final volume of 1.5–2 mL and loaded onto an S-100 gel filtration column (Pharmacia, Piscataway, NJ) equilibrated in 40 mM Bis-Tris pH 7.2, 10 mM NaCl, 2 mM CaCl₂ buffer. Fractions containing protein were run on SDS-PAGE, and only those fractions containing a band corresponding to full-length ssKex2 were retained. The fractions were pooled and concentrated to 25 mg/mL and stored at 4 °C.

Crystallization. ssKex2 (25 mg/mL, 40 mM Bis-Tris pH 7.2, 10 mM NaCl, 2 mM CaCl₂) was crystallized by the hanging drop method against 2.1 M NH₄SO₄, 3% DMSO at 25 °C. Crystals were observed to grow over a period of 2–4 weeks from 6 μ L drops containing 4 μ L of protein solution and 2 μ L of well solution.

Crystal Freezing and Derivatization. Crystals of ssKex2 were transferred from the growth drop into a depression plate containing a 10 μ L drop of 50% saturated sodium malonate, pH 7.2. The crystals were allowed to equilibrate for a period of several minutes. Heavy atom derivatized crystals were obtained by transferring to an identical 20 μ L drop containing either 5 mM Hg-PNP or 5 mM K₂OsCl₆. Crystals were allowed to incubate overnight by sealing the depression wells with tape. After incubation, the crystals were back-soaked by transfer to a 10 μ L drop of 50% saturated sodium malonate, pH 7.2 for 15 min and then cryocooled in liquid nitrogen.

Data Collection. Two heavy atom derivatives (Hg-PNP and K₂OsCl₆) and a 2.4 Å resolution native data set were used to solve the structure of ssKex2. All data were collected on cryocooled crystals maintained at 100 K throughout data collection. The native data were collected at 19ID at the Advanced Photon Source (APS) using a SBC2 3 × 3 CCD detector; the Hg-PNP derivative MAD data sets were collected at beamline 8.3.1 at the Advanced Light Source (ALS) using an ADSC CCD detector; and the K₂OsCl₆ derivative data set was collected at beamline 11-1 at the Stanford Synchrotron Light Source (SSRL) using an ADSC Q315 CCD detector. For both the Hg-PNP and the K₂OsCl₆ derivatives, data were collected in inverse beam mode to collect Friedel mates as close in time as possible. All data were integrated and scaled with DENZO and SCALEPACK, respectively (34). See Table 1 for data statistics.

Structure Determination and Refinement. Initial phases were obtained from a 3.1 Å resolution two wavelength MAD data set collected at the Hg LIII edge of Hg-PNP-soaked ssKex2 crystals. SOLVE (35) was used to locate the Hg atoms and calculate phases, which produced experimental electron density maps of relatively poor quality despite the reasonable figure of merit (0.50). Low occupancy of the Hg sites and poor data quality (Table 1) are the likely reasons that these data failed to generate better phases. Density modification using RESOLVE (35) improved the quality of the electron density significantly and allowed the automatic model building feature of RESOLVE to place a partial polyaniline model into the improved electron density map. Phases calculated from the partial polyaniline model were

Table 1: Data and Model Statistics for ssKex2 2.4 Å Structure

dataset	native	K ₂ OsCl ₆	Hg-PNP	Hg-PNP
beam line	APS 19-ID	SSRL 11-1	ALS 8.3.1	ALS 8.3.1
wavelength (Å)	0.91800	0.95369	1.0050	1.1271
space group	P6 ₅ 22	P6 ₅ 22	P6 ₅ 22	P6 ₅ 22
unit cell (Å)	$a = b = 113.8$ $c = 370.2$	$a = b = 113.6$ $c = 363.3$	$a = b = 113.5$ $c = 369.8$	$a = b = 113.5$ $c = 369.8$
resolution limits (Å)	50.0–2.4	50.0–2.8	50.0–2.8	50.0–2.8
unique reflections	55 847	71 939	57 748	44 566
completeness ^a (%)	98.4 (85.2)	99.8 (100)	87.9 (48.3)	67.8 (6.3)
redundancy ^a	14.2	9.7	4.8	3.4
$I/\sigma(I)$ ^a	20.6 (2.0)	14.4 (2.8)	12.5 (1.2)	19.0 (1.2)
R_{merge} ^{a,b}	0.09 (0.48)	0.15 (0.73)	0.09 (0.59)	0.06 (0.54)
FOM solve	NA	0.27	0.50	0.50
FOM resolve	NA	0.53	0.57	0.57
number of sites	NA	5	6	6
molecules/ASU	2			
solvent content (%)	63			
amino acids residues	954			
water molecules	492			
calcium ions	6			
carbohydrate residues (Glc-Nac)	6			
R_{work} ^c (%)	20.9			
R_{free} ^d (%)	24.5			
av B factor	34.5			
Luzzati coordinate error (Å)	0.30			
bond length RMSD (Å)	0.01			
bond angle RMSD (deg)	1.30			

^a Values in parentheses represent statistics for data in the highest resolution shells. The highest resolution shell comprises data in the range of 2.5–2.4 Å for the native data and 2.9–2.8 Å for the osmium and mercury datasets. ^b $R_{\text{merge}} = \sum_{hkl} \sum_i |I_{hkl} - \langle I_{hkl} \rangle| / \sum_i I_{hkl}$ where i is the i th observation of a reflection with index hkl , and the angle brackets indicate an average over all i observations. ^c $R_{\text{work}} = \sum_{hkl} |F_{hkl}^c - F_{hkl}^o| / \sum_{hkl} F_{hkl}^o$ where F_{hkl}^c is the magnitude of the calculated structure factor with index hkl , and F_{hkl}^o is the magnitude of the observed structure factor with index hkl . ^d R_{free} was calculated as R_{work} , where the F_{hkl}^o were taken from a set of 3033 reflections (5% of the data) that were not included in the refinement (39).

used to locate the five Os sites in $F_o(\text{Os}) - F_o(\text{native})$ difference Fourier maps in a higher resolution K₂OsCl₆ derivative. After SIRAS phasing with SOLVE and density modification with RESOLVE, a more extensive polyaniline model was built. A homology model, based upon the X-ray structures of subtilisin BPN' and Carlsberg (36), of the subtilisin domain of Kex2 (M. Levitt and R. S. Fuller, unpublished data) was superimposed on the partial polyaniline model by least squares superposition in O (37). The model was used as a guide for rebuilding in O and allowed nearly all of the backbone to be traced, as well as facilitating putative side chain assignment and modeling in the clearer portions of the electron density. In addition, the NCS operators relating the two molecules in the asymmetric unit could now be determined. Phase probability distributions were calculated from the improved partial model and were combined with experimental SIRAS phase probability distributions in SIGMAA (38). The combined phase information and the NCS operators were used for a final cycle of density modification in RESOLVE, which produced a readily interpretable electron density map and allowed the remainder of the protein to be built.

The resulting model of ssKex2 was refined against the 2.4 Å resolution native data in CNS. All data were used in the refinement and a bulk solvent correction, and anisotropic scale factors were applied to the data. Throughout the refinement, extremely tight NCS restraints were applied (excluding residues 363–369, 428, 473, 489–494, 533, 542, and 596–599), as tests with lower NCS restraint weights did not result in a significant improvement in the R_{free} value (39). Phase information obtained from the final cycle of density modification was included in the early stages of

refinement via the use of a maximum likelihood target function that incorporated Hendrickson–Lattman coefficients (MLHL). Later stages of refinement were performed against a maximum likelihood target based on amplitudes alone (MLF). Initial simulated annealing torsion angle refinement (40) was followed by several alternating cycles of manual model adjustment in O, coordinate minimization, and individual B factor refinement. For the initial cycles of model adjustment and rebuilding, prime and switch density modification as implemented in RESOLVE was used to minimize the influence of model bias on the electron density maps. A total of 492 waters, six Ca²⁺, and six sugars were added to the model near the end of refinement, in addition to the peptide boronic acid inhibitor. The final model refined to an R of 20.9% ($R_{\text{free}} = 24.5\%$ for a test set of 5% of randomly chosen reflections). See Table 1 for final model statistics.

RESULTS AND DISCUSSION

Overall Structure. The structure of ssKex2 was solved to 2.4 Å resolution by the SIRAS method utilizing a single osmium derivative. The final model of ssKex2 comprises residues 123–599, lacking the first 14 N-terminal and last 17 C-terminal residues of ssKex2 (data statistics are summarized in Table 1). While there is no discontinuity in the electron density, several loop regions display poor electron density for side chains (181–184, 199–201). This is consistent with the high B factors exhibited by residues in these regions. The refined model has good stereochemistry: 84.2, 14.5, and 1.3% of main chain ϕ/ψ angles are in core, allowed and generously allowed regions as calculated using PROCHECK (41). In solution, ssKex2 behaves as a functional monomer based upon its migration on an S-100 gel-

filtration column (data not shown). Crystals of ssKex2 were found to contain two molecules in the ASU related by NCS. Crystal contacts between molecules A and B occlude the active site of molecule A in the lattice. Because the two molecules are structurally identical, the remainder of the discussion will focus solely upon one monomer. As shown in Figure 1, ssKex2 is comprised of two domains, an N-terminal subtilisin-like protease domain (residues 123–457) and a C-terminal P-domain (464–599) that are connected by an extended tether (458–463). The protease domain is structurally similar to subtilisin and related degradative subtilisin family members. Threading the primary sequence of Kex2 against all available protein sequences gave highest scores using thermite and proteinase K. A superposition of subtilisin (PDB: 1AVT (42)) or thermite (PDB: 3TEC (43)) onto ssKex2 results in an overall C α -RMSD of 1.5 and 1.6 Å, respectively, indicating that the entire family of subtilisin-like proteases are very similar to the protease domain of Kex2 in their global fold. Nevertheless, none of the available structures of the degradative subtilisins could be used successfully as a molecular replacement model (data not shown). Both thermite and ssKex2 contain three Ca²⁺ ion binding sites. Of these, the two higher affinity thermite/subtilisin Ca²⁺ ion binding sites (sites two and three in ssKex2) are conserved. The third Ca²⁺ ion binding site in ssKex2 (site 1) is not present in any degradative subtilisin-like structures, and none of the coordinating ligands to the Ca²⁺ ion are conserved between Kex2 and thermite/subtilisin. In addition, the loop (275–283) that forms one face of the site 1 Ca²⁺ ion binding pocket is an insertion conserved in all members of the pro-protein processing protease family but is absent in thermite (Figure 3). Its absence in the degradative protease structures results in a region that is completely open to solvent. The loop forming the opposite face of the binding site is present in the degradative protease structures; however, in Kex2 it appears to be shifted and stabilized by one of the two disulfide bonds (C322–C352) present in the enzyme. The second disulfide (C230–C377) seems to provide similar stabilization to the region forming part of the binding site of one of the structural Ca²⁺ ions, with the backbone carbonyl of C230 contributing a ligand to the metal.

The classic serine protease catalytic triad (D175, H213, and S385) is present in ssKex2, while N314 is correctly positioned to serve as part of the oxyanion hole (Figure 4A). These observations confirm conclusions based upon mutational and functional analysis (30). Cysteine 217, one of the three free thiols in ssKex2 conserved in all of the Kex2/furin proteases and also in thermite and proteinase K but not in subtilisin BPN', is located at a distance of 4.4 Å from both D175 and H213 and may provide a mechanism for the observed inactivation of Kex2 by cysteine modification (44) (Figure 4). The exposed disulfide (C322–C352) that stabilizes a portion of the site 1 Ca²⁺ ion binding site may also be involved in the observation that ssKex2 is inactivated by DTT but is resistant to inactivation by 25 mM β -mercaptoethanol (44).

In addition to the aforementioned features, ssKex2 contains two N-linked glycosylation sites. One site is located in the protease domain at N163 and one in the P-domain at N480 (Figure 1). The maps were of suitable quality to allow for the modeling of one Glc-Nac attached to N163, while two Glc-Nac residues could be modeled at N480. These sites

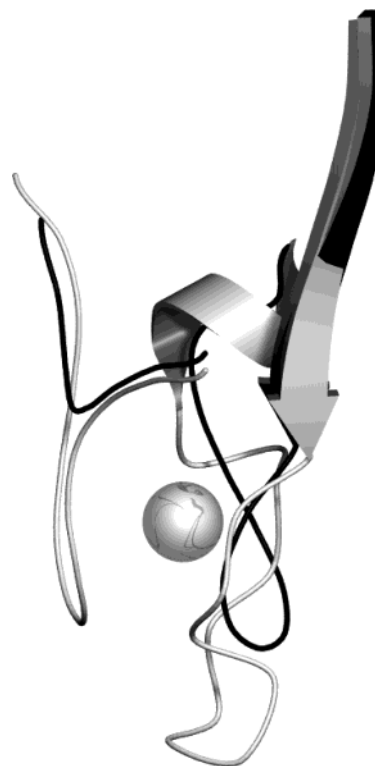


FIGURE 3: Site 1 Ca²⁺ ion binding site. Backbone (C α) trace of the superpositioning results of ssKex2 (gray) and thermite (black) (PDB: 3TEC) in the region of the site 1 Ca²⁺ ion binding site show the absence of the loop (Kex2: 275–283) forming one face of the binding site in thermite and the repositioning of the residues forming the other binding face (Kex2: 312–325).

represent two of the three consensus sequences for N-linked glycosylation present in Kex2 and are consistent with mutational analysis that demonstrates these two sites are the only glycosylated sites *in vivo* (T. Komiyama, A. Neeaga, and R. S. Fuller, personal communication).

P-Domain Structure. The P-domain has a novel jelly roll like fold comprised primarily of two antiparallel β -sheets with a single helix, spanning residues 493–501, connecting strands three and four (Figure 1). While previous theoretical modeling of this domain correctly identified the secondary structural elements, it was incorrect in their lengths and absolute positions in the primary sequence except for strand 5 and failed to identify strand 1 (45). It was predicted that the helical segment would be slightly longer and shifted in position in the primary sequence (499–509); however, a portion of this predicted helix is actually β -strand 4 (502–516) in the crystal structure. The β -strands were also predicted to be considerably shorter than observed in the crystal structure. In soluble proteins, strands in β -sheets typically fall in the range of 3–8 residues in length (46). The β -strands of the P-domain are on the long side of this range (6–9 residues), with the exception of strand 5 (4 residues). The β -strands (4 and 9) from residues 502–516 and 579–594 are quite long, consisting of 15 and 16 residues, respectively. The model structure was also incorrect in the relative three-dimensional positioning of the strands. It correctly identified that the domain would consist of two sheets; however, it suggested one sheet would be formed by strands 3, 2, 5, and 8 and the other by strands 4, 9, 6, and 7 (Kex2 numbering) (45). The two β -sheets in the P-domain are actually comprised of strands 3, 8, 5, and 6 and 2, 9, 4,

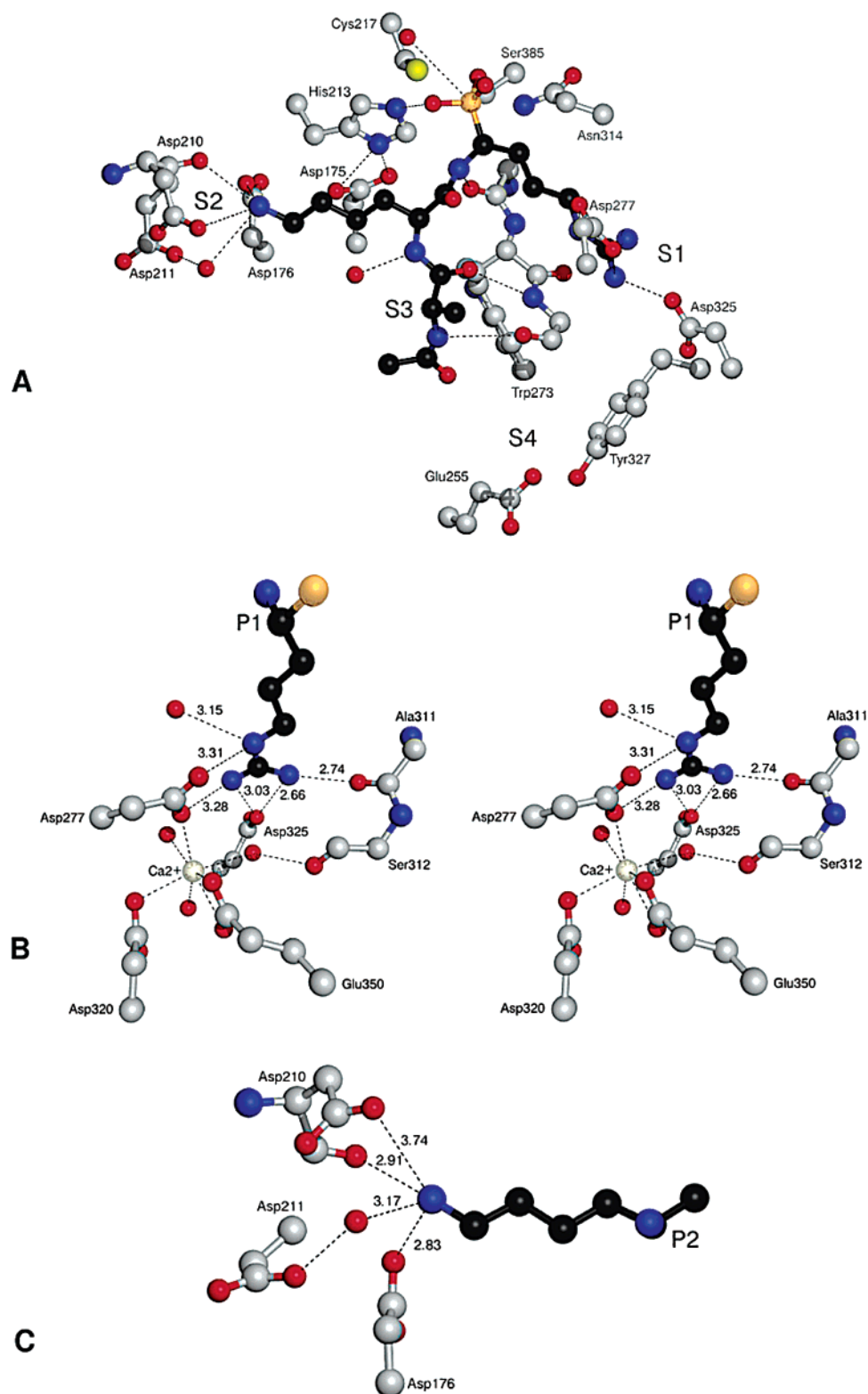


FIGURE 4: ssKex2 active site and S₁–S₄ subsites. Those amino acids making contacts with the inhibitor and Ca²⁺ ion are shown, and their distances are indicated. The atoms are colored according to atom type, and the boron atom is rendered in gold. (A) Overall view of the subsite architecture illustrating the arrangement of the subsites. The catalytic triad D175, H213, and S385; the oxy-anion hole N314; the acyl-enzyme hydrolyzing water molecule; and the reactive C217 are also shown. (B) Close up view of the S₁ and (C) S₂ binding sites illustrating with distances those residues important for substrate and calcium ion recognition and binding.

and 7. A second smaller sheet domain is formed by strands 1, 9, and 4 (Figure 1).

The structure of the ssKex2 P-domain was compared with all other unique folds at the Dali server (<http://www2.ebi.ac.uk/dali/>; (47)). This comparison fails to find any structures with

significant structural identity to the Kex2 P-domain fold (data not shown). While many β -barrel like structures are identified by this search (i.e., viral envelope proteins and carbohydrate binding proteins), none have scores high enough to indicate more than remote structural similarity. The most similar folds

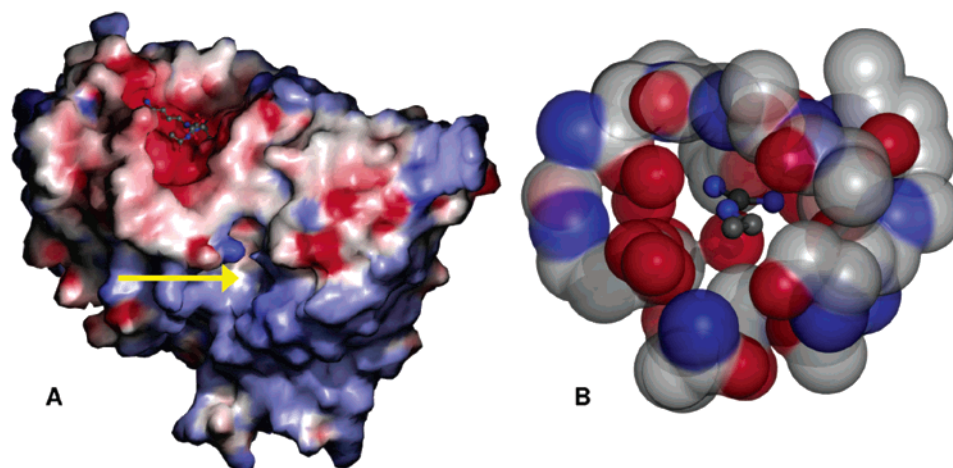


FIGURE 5: Electrostatic surface representation of ssKex2. (A) The electrostatic surface was calculated using the GRASP program (57). Red, blue, and white regions represent negatively charged, positively charged, and neutral regions of the protein, respectively. The arrow indicates the location of the cleft lying between the proteolytic and the P-domains. The protease binding pocket is shown containing the covalently bound inhibitor in a ball-and-stick representation and illustrates the open nature of the S_2 – S_4 binding sites. The P_1 arginine side chain is completely occluded from view by the narrow S_1 binding pocket. (B) The narrow channel forming the P_1 binding site is shown as viewed from the active site looking along the P_1 arginine side chain. Positioned at the bottom of the pocket perpendicular to the plane of the arginine guanidinium group is the carboxylate of D325. The site 1 calcium ion is situated behind D325 and is not visible from this orientation.

have RMSD values greater than 3 Å and Dali scores between 6.0 and 6.8. Therefore, the P-domain of Kex2 has little structural identity to proteins of known structure. It has previously been demonstrated that the P-domain is essential for activity of Kex2 *in vivo* (29). The structure does not rule out a role for the domain in correct processing of mature Kex2 *in vivo*; however, it does suggest that this domain is not directly involved in catalysis. The possibility that interactions between the P-domain and the protease domain stabilize the critical P_1 binding site is discussed below. There does exist a significant cleft between the protease and P-domains as shown in Figure 5A. This cleft is lined predominantly with hydrophobic residues and may represent an extension of the binding pocket from the protease domain or a discrete binding pocket for a yet unidentified ligand. In addition, R540 and R542 are situated in close proximity to the backbone carbonyls of H281 and D278, respectively, which may imply a role for the P-domain in formation of the site 1 Ca^{2+} ion binding site. Further investigation is needed to clarify the functional role of the P-domain in substrate recognition *in vivo* or determine whether it performs an additional function.

Subtilisin-Like Domain: Active Site Calcium Ion Binding Site. The electron density of the active site of ssKex2 clearly shows the presence of the inhibitor, Ac-Ala-Lys-boroArg, covalently bound to S385 (Figure 2). This complex provides significant insight into the selectivity of Kex2 and possibly the entire pro-protein processing protease family. As evidenced by the overall C α -RMSD (1.6 Å) for the protease domain of ssKex2, the active site architecture of ssKex2 is quite similar to that of thermolysin, preserving the locations of the subsites S_1 – S_3 (Figure 4). However, a Ca^{2+} ion binding site is located at the bottom of S_1 that is not observed in the degradative subtilases (Figure 4B). Because of the presence of this ion binding site, there is a significant perturbation of the residues in this region relative to thermolysin (Figure 3). This site appears to play a key role in recognizing arginine at P_1 (Figure 4B). Because of the

increase in the number of acidic residues in the primary sequence of Kex2 when compared to that of the degradative subtilisins, and the fact that S_1 and S_2 preferentially bind substrates with basic residues at P_1 and P_2 , it was predicted that these sites would have a concentration of acidic residues and localized net negative charge (48, 49). While this seems to be the case (Figures 4A,B and 5A), at S_1 the large number of acidic residues present are partially neutralized by the interaction with the active site Ca^{2+} ion. Figure 4B shows that the Ca^{2+} ion is coordinated by D276, D320, and a bidentate interaction with the carboxylate of D350 in addition to three water molecules, forming a seven-coordinate geometry about the metal. Previous experiments have demonstrated a strict dependence of Kex2 activity on Ca^{2+} (3, 44). Of the metals tested (Ca^{2+} , Mg^{2+} , Mn^{2+} , Co^{3+} , and Fe^{3+}), only reconstitution with Ca^{2+} allowed for recovery of any activity (3). Soaks of Ca^{2+} -Kex2 crystals in the presence of 5 mM Tb^{3+} , Lu^{3+} , Eu^{3+} , Ho^{3+} , Tb^{3+} , or Yb^{3+} salts for 24 h failed to achieve crystals with any measurable amount of the lanthanides bound based upon anomalous difference Fourier maps (data not shown). This would further argue that all three calcium sites have very high affinities for calcium. On the basis of these observations, it would appear that Kex2 has evolved a very selective site tuned to the binding of Ca^{2+} ions, and this selectivity is derived from both the size of the binding site as well as the nature of the coordinating ligands, which are exclusively hard oxygen ligands, as is preferred by Ca^{2+} . This discrete selectivity may suggest a role for calcium in regulation of Kex2 activity *in vivo*.

S_1 Binding Site. Previous biochemical studies utilizing substrates with arginine, lysine, ornithine, and citrulline at P_1 have suggested that Kex2 selects for arginine at this position based upon the size and shape of the side chain, the terminal positive charge, and the hydrogen-bonding characteristics of the extended side chain (14, 19, 28). The interaction of the boronate inhibitor with ssKex is shown in Figures 2 and 4. As illustrated, Kex2 interacts with the

arginine side chain at P₁ through a complex series of interactions. The S₁ binding site is a deep narrow pocket (Figure 5B) suggesting that, like S₂, large residues would be excluded (*vide infra*). The site 1 calcium ion is located at the bottom of the S₁ pocket, which suggests that Ca²⁺ plays a role in substrate recognition at P₁. An electrostatic interaction between the terminal positive charge on arginine occurs with D277 and D325, while the calcium plays an indirect role in this interaction through direct coordination of the other carboxylate of D277 (Figure 4B). Therefore, the calcium ion, while not directly interacting with P₁, functions to position D277 to interact with the P₁ arginine side chain. In addition to the electrostatic interaction mediated by the calcium ion, a hydrogen bonding interaction exists between the carboxylate oxygen of D277 and the ϵ -nitrogen of the P₁ arginine (Figure 4B). A third interaction between Kex2 and the P₁ arginine is manifest in an additional hydrogen bond between one of the terminal amines of the arginine side chain and the backbone carbonyl oxygen of A311. Again, this interaction seems to indirectly involve the calcium ion at site 1 through a second sphere coordination of the backbone carbonyl of S312 via an intervening water molecule, helping to position this loop (Figure 4B). The aliphatic portion of the arginine side chain has very little interaction with the S₁ pocket. All of the residues lining the pocket have their side chains oriented outward with the exception of A311 whose methyl group is 4.1 Å from the δ -methylene of the P₁-arginine. On the basis of these points of interaction, Kex2 has evolved a very stringent method of selecting those substrates that have arginine at P₁. Even with the substitution of lysine for arginine at P₁, two of the three interactions are lost. The loss of these two interactions could easily account for the severely hindered ability of Kex2 to proteolyze these substrates, resulting in the observed 100–10 000-fold drop in $k_{\text{cat}}/K_{\text{M}}$ by substituting lysine for arginine at P₁ (28, 50).

S₂ Binding Site. In contrast to the S₁ subsite, the S₂ subsite utilizes a set of much simpler interactions. S₂, while selecting against large amino acids at P₂ (19), is a relatively open pocket (Figures 4 and 5A). The location of H213 (the catalytic histidine) at the mouth of the site may act as a steric filter against bulky side chains. The interaction of substrates with S₂ appears to be primarily electrostatic in nature. In the case of the bound inhibitor, this occurs through direct interaction between the ϵ -amino group of lysine with the carboxylate groups of D176 and D210 (Figure 4C). In addition to these direct interactions, a large negative electrostatic potential is present at S₂ (Figure 5A). The surface of the binding site is lined by D175 and D211 located 4.5–5 Å from the ϵ -nitrogen of the P₂ lysine (Figure 4A). These observations are again consistent with the kinetic studies that show very little distinction between lysine and arginine at P₂ (relative $k_{\text{cat}}/K_{\text{M}}$ = 0.83), with the positive charge at this position providing approximately 3 kcal/mol of transition state stabilization (50).

S₃ Binding Site. Substitute P₃ is an open pocket with no direct interactions between protein and substrate (Figure 4A). This is similar to thermolysin and subtilisin that show no selectivity at P₃ (51). Kinetic analysis of Kex2 has shown no positive selectivity exists at P₃; however, the enzyme may select against aspartate at this position (50). On the basis of the positioning of the alanine side chain of the inhibitor, D276 would lie at a distance of ~4 Å from the carboxylate of a

P₃ aspartate, thereby providing weak electrostatic repulsion that may account for the observed kinetic results. On the basis of this observation, one would expect substrates with P₃ glutamate substitutions to be even more heavily selected against.

S₄ Binding Site. Even though the boronate inhibitor contains no P₄ residue, the superpositioning results with thermolysin in complex with eglin C (43) allow for the tentative determination of the location of the S₄ subsite in Kex2. Kinetic evidence suggests that Kex2 has a dual selectivity for P₄ residues, selecting for aliphatic and basic residues at this position. These same experiments show that the electrostatic interaction at this position is responsible for 2 kcal/mol of transition state stabilization (52). The structural data are consistent with this observation. The mouth of the S₄ pocket is comprised of W273 and Y327 with E255 positioned at the bottom of the pocket to provide the electrostatic interaction with basic residues at P₄ (Figure 4A). While Kex2 is relatively flexible in its selectivity at this position, substitutions for arginine at P₄ essentially abolish activity in Furin (53). Comparisons of the structural determinants of selectivity of this site would be of great interest and await the availability of a structure of Furin.

Additional Enzyme/Inhibitor Interactions. The alignment with thermolysin also shows the location of the P₁' binding site. There appears to be no contacts between ssKex2 and bound substrate at this site and is an open pocket on the enzyme surface adjacent to the P₂ binding site. Since the P₁' and P₂ binding sites are adjacent, very large amino acids at P₁' may interfere with catalysis through perturbation of the interactions between P₂ and S₂ (data not shown). These observations are consistent with kinetic experiments that have shown Kex2 has no selectivity at P₁' but may disfavor bulky amino acids at this position (21).

Backbone interactions between Kex2 and the bound inhibitor are present between the carboxylate oxygens and amide nitrogens of the P₃ alanine and G274 (Figure 4A). These interactions should not depend on the side chain identity at P₁–P₃ and would provide a general mode of recognition and binding for all protein substrates.

The interactions described above most likely manifest themselves in the altered kinetics of nonoptimal P₁–P₄ substrates by both decreasing the apparent affinity of the enzyme for these substrates and by perturbing the binding at S₁ and S₂, thereby interfering with optimal geometry and positioning of the scissile carbonyl for nucleophilic attack by serine 385. The latter conclusion is consistent with the observation that positive interactions at P₄ can compensate partially for substrates with poor P₁ substitutions (52) and with recent studies on chymotrypsin that suggest extended substrate interactions enhance catalysis via a mechanism of distortion (54).

Acyl-Enzyme Hydrolysis. Kinetic evidence has suggested that enzyme deacylation is rate limiting in Kex2 for substrates that contain a P₁ arginine (17, 28) unlike typical serine proteases, including the degradative subtilisins, in which the rate-limiting step has been shown to be acylation (15, 55). It was also proposed that this mechanistic change with Kex2 provides an additional mechanism for selectivity in an environment of incorrect physiological substrates by decreasing the pool of free enzyme (17). In the current structure, a water molecule is positioned in the active site in position to

catalyze the deacylation of the acyl-enzyme intermediate, albeit at a rather long distance (H_2O –boron, 3.6 Å, Figure 4). This water is in a similar position to a water molecule in the structure of subtilisin in complex with D-para-chlorophenyl-acetamido boronic acid (PDB: 1AVT (42)). However, that water molecule is closer to the boron of the derivatized serine (3.3 Å). In the subtilisin structure, the water molecule is located 4.0 Å from the ϵ -N of the catalytic histidine (H64) and appears to hydrogen bond with the δ -sulfur of M222 and the backbone carbonyl of N218. In contrast to the environment of the water molecule in the subtilisin structure, the water molecule in Kex2 would appear to be involved in a network of interactions that could lead to a perturbation of its pK_a . The hydrolyzing water molecule appears to be polarized via interactions with H213 (H_2O –N ϵ 2, 3.8 Å, the catalytic histidine) and H381, the latter interaction occurring through an intervening water molecule (H_2O – H_2O = 3.6 Å; H_2O –N δ 1 = 2.6 Å). H381 is itself polarized by an interaction with E220 (N ϵ 2– H_2O , 2.9 Å). While these polarizing effects would suggest the active site water could function as a better nucleophile, they also lock the water into a position further away from the site of acyl-enzyme hydrolysis. If this water molecule is responsible for deacylation of the enzyme intermediate and the ester carbonyl of the acyl enzyme is positioned where the boron of the inhibitor is located, this increased water–boron distance may provide a basis for the decreased rate of deacylation observed in the kinetic experiments with Kex2.

In conclusion, the reported 2.4 Å resolution crystal structure of the ssKex2–boronate inhibitor complex provides insight into the structural basis by which Kex2 and the dibasic pro-protein processing protease family achieves their high degree of specificity necessary for their diverse physiological functions. The role of the P-domain remains unclear. Clearly, the P-domain does not actively participate in catalysis; however, it may function in substrate recognition through extended contacts via a deep hydrophobic cleft and be involved in formation of the P₁ binding site.

NOTE ADDED IN PROOF

Lipkind et al. (36) in their modeling of the structure of PC1 correctly predicted the interaction of D320 (D325 in Kex2) with the P1 arginine side chain.

ACKNOWLEDGMENT

The authors wish to thank Drs. Charlie Brenner and Geoff Stamper for their contributions early in the project, Dr. Tomoko Komiyama for her general interest and helpful discussions, Tom Alber and James Holden at ALS beamline 8.3.1 for collection of the mercury-derivative dataset, and Sergey Korolev at APS beamline 19-ID for his assistance with collection of the native dataset. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U.S. Department of Energy under Contract DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory. Use of the Argonne National Laboratory Structural Biology Center beamlines at the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Biological and Environmental Research under Contract W-31-109-ENG-38. Portions of this research were carried

out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences.

REFERENCES

- Julius, D., Brake, A., Blair, L., Kunisawa, R., and Thorner, J. (1984) *Cell* 37, 1075–89.
- Fuller, R. S., Brake, A. J., and Thorner, J. (1989) *Science* 246, 482–6.
- Fuller, R. S., Brake, A., and Thorner, J. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86, 1434–8.
- Fuller, R. S., Sterne, R. E., and Thorner, J. (1988) *Annu. Rev. Physiol.* 50, 345–62.
- Komano, H., and Fuller, R. S. (1995) *Proc. Natl. Acad. Sci. U.S.A.* 92, 10752–6.
- Martin, C., and Young, R. A. (1989) *Mol. Cell Biol.* 9, 2341–9.
- Rogers, D. T., Saville, D., and Bussey, H. (1979) *Biochem. Biophys. Res. Commun.* 90, 187–93.
- Cappellaro, C., Mersa, V., and Tanner, W. (1998) *J. Bacteriol.* 180, 5030–7.
- Newport, G., Kuo, A., Flattery, A., Gill, C., Blake, J. J., Kurtz, M. B., Abruzzo, G. K., and Agabian, N. (2003) *J. Biol. Chem.* 278, 1713–20.
- Bader, O., Schaller, M., Klein, S., Kukula, J., Haack, K., Muhlschlegel, F., Korting, H. C., Schafer, W., and Hube, B. (2001) *Mol. Microbiol.* 41, 1431–44.
- Zhou, A., Webb, G., Zhu, X., and Steiner, D. F. (1999) *J. Biol. Chem.* 274, 20745–8.
- Smeekens, S. P. (1993) *Bio/Technol.* 11, 182–6.
- Siezen, R. J., and Leunissen, J. A. (1997) *Protein Sci.* 6, 501–23.
- Rockwell, N. C., Krysan, D. J., Komiyama, T., and Fuller, R. S. (2002) *Chem. Rev.* 102, 4525–48.
- Perona, J. J., and Craik, C. S. (1995) *Protein Sci.* 4, 337–60.
- Brenner, C., Bevan, A., and Fuller, R. S. (1993) *Curr. Biol.* 3, 498–506.
- Rockwell, N. C., and Fuller, R. S. (2001) *Biochemistry* 40, 3657–65.
- Schechter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* 27, 157–62.
- Rockwell, N. C., Wang, G. T., Krafft, G. A., and Fuller, R. S. (1997) *Biochemistry* 36, 1912–7.
- Krysan, D. J., Rockwell, N. C., and Fuller, R. S. (1999) *J. Biol. Chem.* 274, 23229–34.
- Rholam, M., Brakch, N., Germain, D., Thomas, D. Y., Fahy, C., Boussetta, H., Boileau, G., and Cohen, P. (1995) *Eur. J. Biochem.* 227, 707–14.
- Smith, C. A., Toogood, H. S., Baker, H. M., Daniel, R. M., and Baker, E. N. (1999) *J. Mol. Biol.* 294, 1027–40.
- Wilcox, C. A., Redding, K., Wright, R., and Fuller, R. S. (1992) *Mol. Biol. Cell* 3, 1353–71.
- Power, S. D., Adams, R. M., and Wells, J. A. (1986) *Proc. Natl. Acad. Sci. U.S.A.* 83, 3096–100.
- Zhu, X. L., Ohta, Y., Jordan, F., and Inouye, M. (1989) *Nature* 339, 483–4.
- Germain, D., Dumas, F., Vernet, T., Bourbonnais, Y., Thomas, D. Y., and Boileau, G. (1992) *FEBS Lett.* 299, 283–6.
- Wilcox, C. A., and Fuller, R. S. (1991) *J. Cell Biol.* 115, 297–307.
- Brenner, C., and Fuller, R. S. (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89, 922–6.
- Gluschankof, P., and Fuller, R. S. (1994) *EMBO J.* 13, 2280–8.
- Brenner, C., Bevan, A., and Fuller, R. S. (1994) *Methods Enzymol.* 244, 152–67.
- Jagannathan, S., Forsyth, T. P., and Kettner, C. A. (2001) *J. Org. Chem.* 66, 6375–80.
- McPherson, A. (2001) *Protein Sci.* 10, 418–22.

33. Stamper, G. F. (2000) in *Graduate Program in Biophysics and Structural Biology*, pp xiii, 82, Brandeis University, Waltham, MA.
34. Otwinowski, Z., and Minor, W. (1997) in *Methods Enzymology* (Carter, C. W., Jr., and Sweet, R. M., Eds.) pp 307–326, Academic Press, New York.
35. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. D* 55 (Pt 4), 849–61.
36. Lipkind, G., Gong, Q., and Steiner, D. F. (1995) *J. Biol. Chem.* 270, 13277–84.
37. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A* 47 (Pt 2), 110–9.
38. Read, R. J. (1986) *Acta Crystallogr. Sec. A* 42, 140–9.
39. Brunger, A. T. (1992) *Nature* 355, 472–5.
40. Rice, L. M., and Brunger, A. T. (1994) *Proteins* 19, 277–90.
41. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* 26, 283–91.
42. Stoll, V. S., Eger, B. T., Hynes, R. C., Martichonok, V., Jones, J. B., and Pai, E. F. (1998) *Biochemistry* 37, 451–62.
43. Gros, P., Kalk, K. H., and Hol, W. G. (1991) *J. Biol. Chem.* 266, 2953–61.
44. Fuller, R. S., Brake, A., and Thorner, J. (1986) in *Microbiology* (Levine, L., Ed.) pp 273–8, American Society for Microbiology, Washington, D.C.
45. Lipkind, G., Zhou, A., and Steiner, D. F. (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95, 7310–5.
46. Shulz, G. E., and S., R. H. (1979) *Principles of Protein Structure*, Springer, New York.
47. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* 233, 123–38.
48. van de Ven, W. J., Voorberg, J., Fontijn, R., Pannekoek, H., van den Ouweland, A. M., van Duijnhoven, H. L., Roebroek, A. J., and Siezen, R. J. (1990) *Mol. Biol. Rep.* 14, 265–75.
49. Siezen, R. J., de Vos, W. M., Leunissen, J. A., and Dijkstra, B. W. (1991) *Protein Eng.* 4, 719–37.
50. Rockwell, N. C., Wang, G. T., Kraff, G. A., and Fuller, R. S. (1997) *Biochemistry* 36, 1912–7.
51. Gron, H., Meldal, M., and Breddam, K. (1992) *Biochemistry* 31, 6011–8.
52. Rockwell, N. C., and Fuller, R. S. (1998) *Biochemistry* 37, 3386–91.
53. Molloy, S. S., Bresnahan, P. A., Leppla, S. H., Klimpel, K. R., and Thomas, G. (1992) *J. Biol. Chem.* 267, 16396–402.
54. Case, A., and Stein, R. L. (2003) *Biochemistry* 42, 335–48.
55. Philipp, M., and Bender, M. L. (1983) *Mol. Cell. Biochem.* 51, 5–32.
56. Fenn, T. D., Ringe, D., and Petsko, G. A. (2003) *J. Appl. Crystallogr.*, in press.
57. Nicholls, A., Bharadwaj, R., and Honig, B. (1993) *Biophys. J.* 64, A166–A166.

BI034434T