Crystal structure of a bacterial signal peptidase apo-enzyme.

IMPLICATIONS FOR SIGNAL PEPTIDE BINDING AND THE SER-LYS DYAD MECHANISM.

Mark Paetzel‡, Ross E. Dalbey¶, and Natalie C. J. Strynadka‡

‡Department of Biochemistry and Molecular Biology, University of British Columbia, 2146 Health Sciences Mall, Vancouver, BC, V6T 1Z3

¶Department of Chemistry, The Ohio State University, Columbus, Ohio 43210

RUNNING TITLE: Crystal structure of E. coli type I signal peptidase

Corresponding author:
Dr. Natalie C. J. Strynadka
University of British Columbia
Department of Biochemistry and Molecular Biology
Faculty of Medicine
2146 Health Sciences Mall
Vancouver, B.C. Canada V6T 1Z3
Tel.: 604-822-8032 (lab)
    604-822-0789 (office)
Fax.: 604-822-5227
email: natalie@byron.biochem.ubc.ca
SUMMARY

We report here the x-ray crystal structure of a soluble catalytically active fragment of the *Escherichia coli* type I signal peptidase (SPase Δ2-75) in the absence of inhibitor or substrate (apo-enzyme). The structure was solved by molecular replacement and refined to 2.4 Å resolution in a different space group (P4₁2₁2) from that of the previously published acyl-enzyme inhibitor-bound structure (P2₁2₁2) (Paetzel, M., Dalbey, R.E., and Strynadka, N.C.J. (1998) *Nature* 396, 186-190). A comparison with the acyl-enzyme structure shows significant side-chain and main-chain differences in the binding-site and active-site regions which result in a smaller S1 binding pocket in the apo-enzyme. The apo-enzyme structure is consistent with SPase utilizing an unusual oxyanion hole containing one side-chain hydroxyl hydrogen (Ser 88 OγH) and one main-chain amide hydrogen (Ser 90 NH). Analysis of the apo-enzyme active-site reveals a potential deacylating water that was displaced by the inhibitor. It has been proposed that SPase utilizes a Ser-Lys dyad mechanism in the cleavage reaction. A similar mechanism has been proposed for the LexA family of proteases. A structural comparison of SPase and members of the LexA family of proteases reveals a difference in the side-chain orientation for the general-base lysine, both of which are stabilized by an adjacent hydroxyl group. To gain insight into how signal peptidase recognizes its substrates, we have modeled a signal peptide into the binding-site of SPase. The model is built based on the recently solved crystal structure of the analogous enzyme LexA (Luo, Y. et al., (2001) *Cell* 106, 585-594) with its bound cleavage-site region.
INTRODUCTION

Type I signal (leader) peptidase (SPase, EC 3.4.21.89) is the membrane bound serine endoprotease responsible for cleavage of the amino-terminal signal (or leader) peptide extensions from secretory proteins and some membrane proteins (for recent reviews see references (1) and (2)). The type I SPase from *Escherichia coli* has served as the model gram-negative SPase and is the most thoroughly characterized SPase to date. It has been cloned (3), sequenced (4), overexpressed (5), purified (4,6,7), and kinetically (8), and recently structurally (9) characterized. Site-directed mutagenesis (8,10) and chemical modification (8,11) studies are consistent with SPase utilizing a Ser-Lys dyad mechanism whereby Ser 90 serves as the nucleophile and Lys 145 serves as the general-base. *E. coli* SPase (323 amino acids, 35,988 Da, pI 6.9) contains two amino-terminal transmembrane segments (residues 4-28, and 58-76), a small cytoplasmic region (residues 29-58) and a carboxy-terminal periplasmic catalytic region (residues 77-323). A catalytically active fragment of SPase (SPase Δ2-75) which lacks the two transmembrane segments and the cytoplasmic domain has been cloned, over-expressed, characterized (12,13) and crystallized (14). The Δ2-75 construct required detergent or lipid for optimal activity (13) and crystallization (14). The crystal structure of SPase Δ2-75 in complex with a covalently-bound β-lactam based inhibitor revealed directly that Ser 90 was the nucleophile and that the position of the Nζ of Lys 145 was consistent with a role as the general-base (9) in the catalytic mechanism. The shallow hydrophobic pockets observed adjacent to the catalytic residues helped to explain the Ala-X-Ala substrate specificity. The structure also revealed that SPase is an unusual serine protease in that it attacks the
substrate scissile amide carbonyl from the *si*-face rather than the *re*-face as seen in most serine proteases (15,16).

An important component of the catalytic machinery in serine proteases is the oxyanion-hole that works by neutralizing the developing negative charge on the scissile carbonyl oxygen during the formation of the tetrahedral intermediates. Typically oxyanion holes are formed by two main-chain amide hydrogens that serve as hydrogen bond donors to the developing oxyanion (17). The inhibitor bound structure of *E. coli* SPase (9) revealed only one hydrogen bond donor (Ser 90 NH) to the inhibitors acyl-carbonyl group. Although there was no other potential main-chain hydrogen bond donor it was suggested that the side-chain hydroxyl group from Ser 88 could fulfill the role if it were not blocked by the bulky thiazolidine ring of the inhibitor. Here we present crystallographic evidence which, in agreement with recent mutagenesis data (18), strongly suggests the hydroxyl group of Ser 88 does contribute to the oxyanion hole in the Gram-negative signal peptidases. A detailed analysis of the active-site of SPase Δ2-75 in the absence of inhibitor or substrate (apo-enzyme) reveals significant differences from the inhibitor-bound structure and reveals a strong candidate for the key deacylating water in the proposed mechanism. We also present a structural comparison with other Ser-Lys proteases from the LexA family of proteases. A signal peptide is modeled into the binding-sites of SPase based on the structure of the bound cleavage-site region in LexA.
EXPERIMENTAL PROCEDURES

Materials - The SPase ∆2-75 protein (relative molecular mass (Mr) 27,952 by electrospray ionization mass spectrometry analysis (12), 249 amino acid residues, pI 5.6 by isoelectric focusing analysis (13)) was expressed and purified as described previously (14). The protein (10 mg/ml) was solubilized in 20 mM Tris-HCl pH 7.4 and 0.5% of the detergent Triton X-100.

Crystallization - Subtle adjustments to the previously published crystallization conditions (14) resulted in the formation of a new crystal form of ∆2-75. The crystals of apo-enzyme SPase ∆2-75 were grown by sitting drop vapor diffusion with a reservoir consisting of 0.70 M ammonium dihydrogen phosphate, 0.1 M sodium citrate pH 5.4 and a drop consisting of 4 µl of protein and 8 µl of reservoir solution. The crystals formed in drops that had been left undisturbed for six weeks at a consistent temperature of 22 ºC. The crystals belong to the tetragonal space group P4₁2₁2 with unit cell dimensions of a = 112.4 Å, b = 112.4 Å, c = 198.7 Å. Given the unit cell dimensions and the molecular mass, the specific volume (Vₘ) (19) is 2.78 Å³/Da for four molecules in the asymmetric unit. The x-ray diffraction intensities were measured at 100 K on beamline BL7-1 (Stanford Synchrotron Radiation Laboratory) at a wavelength of 1.08000 Å. The data were processed using the programs DENZO and SCALEPACK (20).

Phasing, Model Building, Refinement and Structural Analysis - A molecular replacement solution was found using the program AmoRe (21) and the atomic coordinates of the previously publish crystal structure of the SPase ∆2-75 inhibitor complex (PDB =1B12). Rebuilding of some of the flexible regions was required due to differences resulting from crystal packing. Model building and analysis was performed...
with the program XFIT within the suite XTALVIEW (22). Refinement of the structure was carried out using the program CNS (23). The secondary structural analysis was performed with the program PROMOTIF (24). The utilities Lsq_explicit, Lsq_improve and Lsq_molecule within the program O (25) were used to superimpose molecules. The binding pocket analysis was performed using the CAST server (26) [http://cast. engr.uic.edu/cast/]

**Modeling of a signal peptide bound to SPase** - A model of a signal peptide bound to the binding-sites of SPase was prepared by using the structure of the covalently bound penem inhibitor complex to guide the position of the P1 Ala residue of the substrate. The P3 Ala position to the P6 substrate position was guided by the structure of the cleavage-site domain of the analogous enzyme LexA (PDB code: 1JHE). The model was prepared using the program O (25). The model was energy-minimized using the program CNS (23).

**Figure Preparation** - Figures 1, 3, 5-7 were prepared using the programs Molscript (27) and Raster3D (28). Figure 4 was prepared using the programs XFIT (22) and Raster3D (28). Figure 2 was prepared using the programs GRASP, UNGRASP (29) and Raster3D (28).

**Accession Numbers** - Atomic coordinates for the apo-enzyme SPase Δ2-75 structure will be deposited with the RCSB Protein Data Bank (30) [http://www.rcsb.org/pdb/] upon acceptance of the manuscript. Atomic coordinates for the SPase Δ2-75 penem inhibitor complex structure (9) are under the accession code 1B12.
RESULTS AND DISCUSSION

The catalytic domain of *E. coli* type I signal peptidase (SPase Δ2-75) has been crystallized in the absence of inhibitor or substrate (apo-enzyme). Earlier attempts to crystallize the apo-enzyme of SPase Δ2-75 resulted in poorly diffracting crystals that belonged to the tetragonal space group P4_{2}2_{1}2 (14). Adjustments to the previously published crystallization conditions have resulted in improved crystals of the apo-enzyme that belong to the tetragonal space group P4_{1}2_{1}2 with diffraction to 2.4 Å resolution (see experimental procedures for details). Typical of many enzyme systems, the apo-enzyme form is apparently less ordered than that of the inhibitor bound form. Orthorhombic (P2_{1}2_{1}2) crystals of the *E. coli* SPase acyl-enzyme inhibitor complex were shown to diffract to beyond 1.9 Å resolution (9). The structure of the apo-enzyme was solved by molecular replacement and refined to 2.4 Å resolution with a R-factor of 0.239 and a R-free value of 0.279 (see table 1 for complete refinement statistics).

**Structural difference between apo-enzyme and acyl-enzyme.**

Both the apo-enzyme structure and the acyl-enzyme inhibitor-bound structure reveal two different paths taken for the N-terminal region (residues 80 - 85), just preceding the nucleophilic Ser 90. In two of the four molecules in the asymmetric unit this region is in a β-strand conformation and in the other two there is a bulge in this region. This difference in the N-terminal region appears to be the result of crystal packing forces. If we compare molecules with similar N-terminal region conformations we see little change in the overall structure upon formation of the inhibitor acyl-enzyme complex. The root mean square deviation (r.m.s.d.) upon superposition of the apo-enzyme structure (molecule A) and the acyl-enzyme inhibitor-bound structure (molecule
A) is 0.558 Å for 234 common Cα atoms. Most of the differences arise from the extended loops and hairpins near the solvent surface (108-124, 170-176, 198-202, and 304-313). Comparing molecules with dissimilar N-terminal region conformations gives a significantly larger r.m.s.d., for example superimposing molecule B from the apo-enzyme structure onto molecule A in the inhibitor-bound acyl-enzyme structure gives an r.m.s.d. of 0.974.

Comparing the apo-enzyme and acyl-enzyme active site residues reveals very little change in the position of the nucleophilic Ser 90 and general-base Lys 145 (Fig. 1,2). There are however significant differences in the substrate-binding pockets. The shallow and hydrophobic substrate-binding pockets of E. coli SPase are consistent with the substrate specificity for small and neutral residues at the P1 and P3 position (P1 and P3 (31) correspond to the -1 and -3 position relative to the cleavage-site) (32). The S1 pocket is made up of nonpolar atoms from residues 86-88, 90-91, 95, and 143-145. The S3 pocket is made up of nonpolar atoms from residues 84, 86, 101, 132, 142 and 144. In the apo-enzyme structure the main-chain of Pro 87 and Ser 88 is positioned such that the S1 binding pocket is narrower than that seen in the acyl-enzyme inhibitor-bound structure (Fig. 2,3). The side chain of Phe 84 has also rotated approximately 80° about \( \chi_1 \), toward the direction of the binding pocket (from -83° in the inhibitor-bound structure to -160° in the apo-enzyme structure). Differences are also seen in the side-chain positions of Ile 144, Leu 95, and Tyr 143 (Fig. 1,2), which also result in a more narrow and shallow binding pocket (Fig.3). Molecular surface analysis of the binding pocket region (26) confirms quantitatively that both the area and the volume of the S1 binding pocket is smaller in the apo-enzyme structure. The S1 binding site in the acyl-enzyme structure is the largest.
pocket on the SPase molecular surface with an area of 156.8 Å^2 and a volume of 224.1 Å^3. The S1 binding site in the apo-enzyme structure is the third largest pocket on the SPase molecular surface with an area of 145.5 Å^2 and a volume of 129.3 Å^3.

Another major difference in the active site is seen in the side chain position of Ser 88, where $\chi_1$ is rotated by ~180° from its position in the acyl-enzyme structure. The position of the Ser 88 hydroxyl group in the apo-enzyme is consistent with its role in the formation of the SPase oxyanion hole and stabilization of the tetrahedral oxyanion intermediate. Paetzel and coworkers had previously proposed that the Ser 88 $\chi_1$ angle in the inhibitor-bound structure was a result of a clash with the inhibitors thiazoladine ring (9) and that in the absence of the bulky inhibitor, the Ser 88 $\beta$-hydroxyl may be able to contribute to the oxyanion stabilization. These structural findings are consistent with recently published mutagenesis and kinetic studies (18) which showed that the S88T mutant maintained near wild-type activity whereas the S88C and S88A mutations resulted in a 740-fold and >2440-fold reduction in activity, respectively. Sequence analysis reveals that SPases from Gram-negative bacteria and those from the eukaryotic endoplasmic reticulum contain a Ser at the 88 position (E. coli numbering). There is some precedent for a serine hydroxyl acting as a component of an oxyanion hole (in lipolytic enzymes such as cutinase (33)), but SPase is the first characterized serine protease to utilize a side-chain hydroxyl in this manner.
Active-site waters.

The crystal structure of the apo-enzyme of SPase Δ2-75 reveals electron density and appropriate coordination for three water molecules near the active-site region conserved in each of the four molecules in the asymmetric unit (Fig. 4). The water designated WAT1 is coordinated by Met 91 NH, Ser 88 O, and Leu 95 O. The water designated WAT2 is coordinated by Ile 144 NH. The water designated WAT3 is coordinated by Lys145 Nζ. WAT1 was also present in the SPase acyl-enzyme inhibitor-bound structure (9). WAT2 AND WAT3 were not present in the acyl-enzyme inhibitor-bound structure as they would have been sterically displaced by the presence of the inhibitor (Fig. 5). This suggests either WAT2 or 3 as the deacylating water, with the long lived stability of the penem inhibitor acyl-enzyme complex a result of displacement of the deacylating water. Inhibitors for TEM-1 β-lactamase have been designed based on this principle (34,35) and analysis of serine protease acyl-enzyme complexes have revealed similar effects (36,37).

For a water to function as the "deacylating" (also called the "catalytic", "hydrolytic" or "nucleophilic") water during the deacylation step of the hydrolytic reaction it needs to be positioned such that it is within hydrogen bonding distance to a general-base for activation as well as at a suitable angle of approach with respect to the carbonyl of the ester intermediate (the so-called Burgi angle of ~107º; (38)). By superimposing the apo-enzyme structure and the inhibitor-bound acyl-enzyme structure we can approximate the distance from WAT 2 and WAT3 to the ester carbonyl carbon as well as the angle subtended by the water oxygen, the ester carbonyl carbon and the carbonyl oxygen (table 2, Fig. 5). WAT2 would most likely be displaced by the binding
of the signal peptide due to its β-sheet style hydrogen bonding interactions with the strand containing Ile 144 which coordinates WAT2 (Fig. 7a). Of the three water molecules seen near the active-site of the SPase apo-enzyme, WAT3 would be the most-likely candidate for a deacylating water. It is approximately equa-distant between Lys 145 Nζ (3.3 Å) and the ester carbonyl carbon (3.8 Å) of the inhibitor (C7), and has an angle of approach to the ester carbonyl carbon (Burgi angle) of 78°.

A structural comparison of Ser/Lys proteases.

Based on our structural evidence, the SPase and the LexA families of proteases belong to the same evolutionary clan (SF) of serine proteases. A structural comparison of SPase and UmuD’ protease revealed that these enzymes share the same protein fold in the catalytic core (39). The crystal structures of two other members of the LexA family of self-cleaving proteases (LexA (40) and λ phage cI (41)) have also been recently solved. The central catalytic domain of SPase can be superimposed on the structures of UmuD’ (42), LexA (40), and cIλ (41) with r.m.s. deviations of 1.54, 1.61, and 1.54 Å for 68, 71, and 74 common Cα atoms respectively (Fig 6a). Analysis of the active-sites of these superimposed structures reveal that these enzymes share a common hydrogen bonding environment for the Lys general-base in that it is hydrogen bonded not only to the Ser nucleophile Oγ but also to another side-chain hydroxyl Oγ (Fig. 6b). The position of the second coordinating hydroxyl Oγ is different in SPase from that of the LexA family of proteases. The Lys 145 Nζ of SPase is hydrogen bonded to Ser 278 Oγ and Ser 90 Oγ. The Lys 156 Nζ of LexA is coordinated by Thr 154 Oγ and Ser 119 Oγ. In SPase the non-nucleophilic Ser 278 Oγ resides on a β-strand that lays adjacent but perpendicular to
the β-strands that contains the general-base Lys 145. The similar hydroxyl Oγ in the LexA-like proteases (Thr 154 in Lex A) resides on the same β-strand as the lysine general-base. The different orientations for the second coordinating hydroxyl Oγ results in slightly different orientations (side-chain χ4 angle) for the ε-amino group of the Lys general-base. For example, the χ4 angle for Lys 145 in the apo-SPase structure is 178.3° (an average for the four molecules in the asu) and the χ4 angle for Lys 97 in the UmuD’ structure (1umu) is -75.5° (an average for the two molecules in the asu). It should be noted that although the Thr residue is conserved in the LexA family, the Thr Oγ in the UmuD’ structure is pointing in the opposite direction to that seen in LexA and cIλ structures.

The LexA family of proteases have an additional hydrogen bond to the Lys general base that SPase does not have. The Lys Nζ is within hydrogen bonding distance to the main-chain oxygen of the residue preceding the general base (Fig. 6b). In the case of LexA for example, the neutral ε-amino group of Lys 156 (the deprotonated state is a requirement for it serving as the general-base) would have two hydrogen bond acceptors (Val 155 O and Thr 154 Oγ) and one hydrogen bond donor (Ser 119 Oγ).

Photosystem II D1 C-terminal processing protease (CtpA) has no sequence or structural similarity to SPase or the LexA family of proteases but utilizes a similar Ser-Lys dyad mechanism. The recently solved crystal structure of CtpA shows a non-nucleophilic hydroxyl-group (Thr 168) within hydrogen bonding distance to the proposed Lys general-base (Lys 397) (43). For a list of the structures used in this comparison see
table 2. The class A β-lactamases which utilize a Ser-Lys dyad mechanism also have a second Ser hydroxyl hydrogen bond to the general-base Lys Nζ (44).

It appears from this structural comparison that the Ser-Lys dyad mechanism requires a third player (Ser/Thr Oγ) for optimal activity, much like the Asp in the Ser-His-Asp catalytic triad mechanism. The second hydroxyl hydrogen-bond to the Lys Nζ in the Ser-Lys dyad proteases possibly functions to help align the general-base Lys to the nucleophilic Ser hydroxyl. Mutagenesis data shows that the SPase Ser 278 is essential for optimal activity (45). The three hydrogen bonding contacts with the general-base Lys in the LexA family of proteases suggests that the Lys Nζ in these enzymes would be very firmly held in place.

**Model of a signal peptide bound to a signal peptidase.**

Recently the crystal structure of the self-cleaving LexA repressor active-site mutant (S119A) was solved with its cleavage-site region bound in the active-site (40). Using the acyl-enzyme inhibitor-bound SPase structure, along with the structure of the LexA cleavage-site, we have modeled the signal peptide from *E. coli* outer membrane protein A (OmpA) bound to the *E. coli* SPase (Fig. 7). The OmpA signal peptide was chosen for this modeling study because *E. coli* SPase shows the best catalytic constants using the pre-protein substrate ProOmpA-nucleaseA (PONA) (46). PONA is a hybrid protein of the OmpA signal peptide attached to the *Staphylococcus aureus* nuclease A. The model shows the signal peptide making a number of β-sheet type of hydrogen bonding interactions with the β-strands (81-85, 142-149) that line the S1 and S3 substrate binding pockets (Fig. 7a). The hydrogen bonds include Gln 20 (P2) O ⋯ Ile 144 NH, Gln 20 (P2) NH ⋯ Asp 142 O, Ala 19 (P3) NH ⋯ Gln 85 O, Thr 17 (P5) O ⋯ Gln 85 NH, Ala
16 NH (P6) ⋯ Pro 83 O, and Gly 14 NH (P8) ⋯ Tyr 81 O. This model suggests possible substrate-enzyme contact points all the way to the P8 position of the signal peptide. The P2, P4 and P5 side-chains are mostly solve exposed, which is consistent with the lack of specificity at these positions. The model does however suggest the side-chains of the P6 and P7 residues may be in position to make significant contact with SPase. Interestingly, the P7 residue (Phe 15) comes in close proximity to Trp 300 which has been shown by site-directed mutagenesis (47) and chemical modification (48) to be important for optimal activity. Many signal peptides contain a Gly or Pro residue at approximately the P6 residue suggesting that this may represent the transition point from the β-strand region (C-region) to the α-helical region (H-region). By extending the signal peptide β-sheet type of interaction to the P6-P8 position, the H-region is presented in close proximity to the presumed position of the transmembrane segments of SPase (Fig. 7b).

**Concluding Remarks.**

There is very little sequence identity among signal peptides except in the cleavage-site region (C-region) where there is a preference for small and neutral residues at the P1 and P3 position. By far the most common residues at these positions is Ala, although Gly and Ser are sometimes observed at the P1 position and Val, Ser and Leu are sometimes seen at the P3 position. By comparing the apo-enzyme and acyl-enzyme structures of SPase we can see that the flexibility seen in the shallow and hydrophobic binding site region is consistent with the relatively broad substrate specificity of SPase. Although the β-lactam style SPase inhibitor mimics the substrates amine bond it is not clear whether the same changes seen at the active-site and binding-site upon formation of the inhibitor acyl-enzyme are reflective of the changes that would be observed with the
substrate acyl-enzyme. Unpublished results from a crystal structure of a peptide-based inhibitor non-covalently bound to SPase shows approximately the same Pro 87 and Ser 88 main-chain position as the acyl-enzyme inhibitor-bound structure. Conversely the peptide-based non-covalently bound inhibitor structure shows a similar Ser 88 $\chi_1$ angle as that in the apo-enzyme structure which would be consistent with SPase utilizes a Ser 88 hydroxyl in its oxyanion hole. (Paetzel et al., 2001, unpublished data).

Structural comparisons with the related enzymes from the LexA family of proteases reveals similarities in the hydrogen bonding network for the general-base Lys yet differences in the orientation of the Lys N$\zeta$ supported by the network. It may be that the Ser-Lys dyad mechanism would be more accurately described as a Ser-Lys-Ser/Thr triad mechanism.

Since SPase is an essential enzyme found at the bacterial membrane surface it has been actively studied as a potential antibacterial target (49). The crystal structure of the apo-enzyme form of *E. coli* SPase and our model of the OmpA signal peptide bound to SPase may be helpful in the rational design of inhibitors and artificial substrates as well as kinetic and mutagenesis experiments. Efforts are now underway to crystallize the full-length *E. coli* SPase bound to a pre-protein substrate.
REFERENCES

FOOTNOTES

* This work was supported by the Burroughs Wellcome Foundation New Investigator Award to NCJS, the Canadian Bacterial Diseases Network of Excellence and the Howard Hughes Medical Institute International Scholar Award to NCJS. M.P. is funded by a Canadian Institute of Health Research post-doctoral fellowship and NCJS by a Canadian Institute of Health Research Scholarship.

The atomic coordinates for the inhibitor-free structure of Δ2-75 (PDB code: XXXX) have been deposited in the Protein Data Bank, Research Collaboration for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

|| To whom correspondence should be addressed: Tel.: 604-822-0789, Fax.: 604-822-5227, E-mail: natalie@byron.biochem.ubc.ca

ACKNOWLEDGMENTS

We thank Dr. Henry Bellamy for allowing us to use beam-line BL1-5 at the Stanford Synchrotron Radiation Laboratory (SSRL). We thank Dr. Yu Luo for helpful discussions on the structure of LexA.

ABBREVIATIONS

The abbreviations used are: SPase, signal peptidase; Δ2-75, the construct of E. coli signal peptidase lacking the residues 2 through 75 which correspond to the transmembrane segments and the cytoplasmic region; asu, crystallographic asymmetric unit; r.m.s.d., root mean-square deviation; pI, isoelectric point; Da, Daltons; UmuD', the UmuD protease with the N-terminus cleaved.
FIGURE LEGENDS

FIG. 1. The active-site of *E. coli* SPase apo-enzyme. A stereographic rendering of the active-site of *E. coli* signal peptidase in the absence of inhibitor or substrate. The regions of the model that are colored red correspond to the boxes of conserved amino acid sequence in SPase.

FIG. 2. Comparison of the active-site residues of *E. coli* SPase with and without a bound inhibitor. The apo-enzyme structure is shown in red and the acyl-enzyme inhibitor-bound structure is shown in black. The inhibitor has been removed for clarity.

FIG. 3. The molecular surface of the *E. coli* SPase binding-site. a. The molecular surface of the acyl-enzyme inhibitor-bound *E. coli* SPase. The inhibitor has been removed for clarity. b. The molecular surface of the apo-enzyme of *E. coli* SPase. The hydrophobic surface is shown in green. The S1 substrate binding pocket is indicated.

FIG. 4. Electron density at the active-site of SPase. A cross-validated sigmaa 2Fo-Fc electron density map contoured at 1.5σ surrounding the active-site residues of *E. coli* SPase apo-enzyme.

FIG. 5. Active-site waters in *E. coli* SPase. Each molecule in the asymmetric unit of the SPase apo-enzyme structure has been superimposed to show the positions of the conserved active-site waters in SPase. The active-site of the penem inhibitor-bound structure is also superimposed on the apo-enzyme structure to show how the penem
would have displaced WAT2 and WAT3. Note that WAT1 is also observed in the 
inhibitor-bound structure. The ester carbonyl carbon (C7) and oxygen (O8), which 
mimics the ester of the acyl-intermediate, are labeled.

FIG. 6. The structural comparison of Ser/Lys proteases. a. A Cα trace of SPase 
(blue) (9) overlapped on the C-terminal protease domains of UmuD' (red) (42), cl-λ 
(yellow) (41), and LexA (green) (40). b. A ball-and-stick rendering of the overlapped 
active-sites of SPase, UmuD', clλ, and LexA. The hydrogen-bond interactions for the Nζ 
of the general-base lysine for signal peptidase (K145) and LexA (K156) are shown.

FIG. 7. A model of the OmpA signal peptide bound to the active-site of the E. coli 
SPase. a. The P8 to P1' residues (Gly 14 - Ala 22) of the OmpA signal peptide have been 
modeled into the substrate binding pockets of E. coli SPase. The model is based on the 
LexA cleavage-site (40). The protein is shown in black stick and the modeled peptide is 
in ball and stick, grey (carbon), blue (nitrogen) and red (oxygen). Potential hydrogen 
bonding interactions between the signal peptide and SPase are shown. b. A ribbon 
diagram of the OmpA signal peptide modeled into the binding-site of E. coli SPase Δ2- 
75. The signal peptide is shown in purple. The positions of the P1, P3 and P7 residues 
are indicated. The Ser 90 and Lys 145 are shown in ball-and-stick. Two phospholipid 
molecules rendered in CPK are positioned next to the complex to give a perspective of 
the thickness of the phospholipid bilayer relative to the signal peptidase/signal peptide 
complex. The phospholipid molecules were taken from a molecular dynamics simulation 
of a bilayer (52). The regions of E.coli SPase that are present in all type I SPases are
shown in red and green. The red regions correspond to the conserved boxes of sequence.
Table 1. Crystallographic data

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>P4₁₂₁₂ (#92)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P4₁₂₁₂ (#92)</td>
</tr>
<tr>
<td>Unit cell dimensions a, b, c (Å)</td>
<td>112.4 x 112.4 x 198.7</td>
</tr>
<tr>
<td>Molecules in a.s.u.</td>
<td>4</td>
</tr>
<tr>
<td>Vm (Å³/Da)</td>
<td>2.78</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.0 - 2.40</td>
</tr>
<tr>
<td>Total observed reflections</td>
<td>277,201</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>49,984</td>
</tr>
<tr>
<td>% possible</td>
<td>98.8 (100.0)</td>
</tr>
<tr>
<td>I / σ(I) (%)</td>
<td>25.4 (5.8)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>4.1 (29.6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues</td>
<td>907</td>
</tr>
<tr>
<td>Atoms</td>
<td>6912</td>
</tr>
<tr>
<td>Waters</td>
<td>258</td>
</tr>
<tr>
<td>R</td>
<td>23.9</td>
</tr>
<tr>
<td>R_free</td>
<td>27.9</td>
</tr>
<tr>
<td>r.m.s. deviations</td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.0066</td>
</tr>
<tr>
<td>Angles (°)</td>
<td>1.4524</td>
</tr>
<tr>
<td>Overall B (Å²) for protein atoms</td>
<td>45.1</td>
</tr>
<tr>
<td>Overall B (Å²) for water atoms</td>
<td>40.9</td>
</tr>
</tbody>
</table>

$R_{\text{merge}} = \sum |I_{o,i} - |I_{\text{ave},i}| / \sum |I_{\text{ave},i}|$, where $I_{\text{ave},i}$ is the average structure factor amplitude of reflection $I$ and $I_{o,i}$ represents the individual measurements of reflection $I$ and its symmetry equivalent reflection.

$R = \sum |F_o - F_c| / \sum F_o$ (on all data 2.4 - 50.0 Å)

$R_{\text{free}} = \sum_{hkl \subset \text{T}} (|F_o| - |F_c|)^2 / \sum_{hkl \subset \text{T}} |F_o|^2$, where $\sum_{hkl \subset \text{T}}$ are reflections belonging to a test set of 10% of the data, and $F_o$ and $F_c$ are the observed and calculated structure factors, respectively.

The data collection statistics in brackets are the values for the highest resolution shell (2.49 to 2.40 Å).
Table 2. Active-site water molecules in the *E. coli* SPase apo-enzyme structure.

<table>
<thead>
<tr>
<th>WAT †</th>
<th>B-factor (Å²)</th>
<th>coordination (Å)</th>
<th>Distance to Lys 145 Nζ (Å)</th>
<th>Distance to penem ester carbonyl carbon C7 (Å)*</th>
<th>Burgi angle (°) *</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27.8</td>
<td>Met 91 NH (2.7)</td>
<td>8.1</td>
<td>5.1</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ser 88 O (2.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Leu 95 O (2.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>37.2</td>
<td>Ile 144 NH (3.3)</td>
<td>4.4</td>
<td>3.0</td>
<td>112</td>
</tr>
<tr>
<td>3</td>
<td>40.5</td>
<td>Lys 145 Nζ (3.2)</td>
<td>3.3</td>
<td>3.8</td>
<td>78</td>
</tr>
</tbody>
</table>

WAT 1 = solvent atoms; 55S, 21S, 30S, and 15S in the PDB file ####.
WAT 2 = solvent atoms 66S, 481S, 17S and 241S in the PDB code ####.
WAT 3 = solvent atoms 75S, 11S, 78S, and 63S in PDB code ####.
† = The values cited are an average of the values observed in the each of the four molecules in the asymmetric unit.
* = Based on superposition of apo-enzyme structure onto acyl-enzyme inhibitor complex structure.

Burgi angle = The angle subtended by the water oxygen, the ester carbonyl carbon and the ester carbonyl oxygen (Owat - Cacyl -Oacyl) (38).
<table>
<thead>
<tr>
<th>Enzymes</th>
<th>PDB</th>
<th>reference</th>
<th>Res. (Å)</th>
<th>R/R_free</th>
<th>Mol/asu</th>
<th>Ser nucleophile</th>
<th>Lys general-base</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPase-acyl Inhibitor</td>
<td>1B12</td>
<td>(9)</td>
<td>1.9</td>
<td>0.220/0.246</td>
<td>4</td>
<td>90</td>
<td>145</td>
</tr>
<tr>
<td>SPase-apo</td>
<td>####</td>
<td>This work</td>
<td>2.4</td>
<td>0.239/0.277</td>
<td>4</td>
<td>90</td>
<td>145</td>
</tr>
<tr>
<td>UmuD' Se</td>
<td>1UMU</td>
<td>(42)</td>
<td>2.5</td>
<td>0.207/0.303</td>
<td>2</td>
<td>60</td>
<td>97</td>
</tr>
<tr>
<td>UmuD' WT</td>
<td>1AY9</td>
<td>(50)</td>
<td>3.0</td>
<td>0.218/0.287</td>
<td>2</td>
<td>60</td>
<td>97</td>
</tr>
<tr>
<td>UmuD' NMR</td>
<td>1I4V</td>
<td>(51)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>60</td>
<td>97</td>
</tr>
<tr>
<td>cI-λ</td>
<td>1F39</td>
<td>(41)</td>
<td>1.9</td>
<td>0.228/0.257</td>
<td>2</td>
<td>149</td>
<td>192</td>
</tr>
<tr>
<td>LexA Δ1-67, S119A</td>
<td>1JHC</td>
<td>(40)</td>
<td>2.0</td>
<td>0.230/0.278</td>
<td>1</td>
<td>119</td>
<td>156</td>
</tr>
<tr>
<td>LexA, S119A</td>
<td>1JHH</td>
<td>(40)</td>
<td>2.1</td>
<td>0.230/0.273</td>
<td>2</td>
<td>119</td>
<td>156</td>
</tr>
<tr>
<td>LexA, G85D</td>
<td>1JHF</td>
<td>(40)</td>
<td>1.8</td>
<td>0.240/0.263</td>
<td>2</td>
<td>119</td>
<td>156</td>
</tr>
<tr>
<td>LexA, Δ1-67, QM</td>
<td>1JHE</td>
<td>(40)</td>
<td>2.5</td>
<td>0.220/0.284</td>
<td>119</td>
<td>156</td>
<td></td>
</tr>
<tr>
<td>CtpA Se-Met/L153M/L210M</td>
<td>1FC6</td>
<td>(43)</td>
<td>1.8</td>
<td>0.168/0.237</td>
<td>1</td>
<td>372</td>
<td>397</td>
</tr>
<tr>
<td>CtpA Native C2A</td>
<td>1FC7</td>
<td>(43)</td>
<td>2.0</td>
<td>0.187/0.267</td>
<td>1</td>
<td>372</td>
<td>397</td>
</tr>
<tr>
<td>CtpA Native C2B</td>
<td>1FC9</td>
<td>(43)</td>
<td>1.9</td>
<td>0.191/0.275</td>
<td>1</td>
<td>372</td>
<td>397</td>
</tr>
<tr>
<td>CtpA Native R32</td>
<td>1FCF</td>
<td>(43)</td>
<td>2.1</td>
<td>0.220/0.318</td>
<td>1</td>
<td>372</td>
<td>397</td>
</tr>
</tbody>
</table>

SPase = E. coli type I signal peptidase
UmuD' Se = UmuD' with Seleno-Methonine in place of the Methionines
UmuD' WT = Wild-type UmuD'
CtpA = Photosystem II D1 C-Terminal Processing Protease
Mol/asu = molecules per asymmetric unit
LexA, Δ1-67, QM = LexA repressor deletion construct which lacks residues 1 through 67 and contains the mutations (L89P, Q92W, E152A, K156A).