doi:10.1016/j.jmb.2012.03.020 J. Mol. Biol. (2012) 419, 347–358

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Crystal Structure of *Bacillus subtilis* Signal Peptide Peptidase A

Sung-Eun Nam, Apollos C. Kim and Mark Paetzel*

Department of Molecular Biology and Biochemistry, Simon Fraser University, South Science Building 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6

Received 12 January 2012; received in revised form 16 March 2012; accepted 26 March 2012
Available online 1 April 2012

Edited by R. Huber

Keywords:
signal peptide peptidase; self-compartmentalized protease; serine/lysine dyad; lysine general base; serine protease

Signal peptide peptidase A (SppA) is a membrane-bound self-compartmentalized serine protease that functions to cleave the remnant signal peptides left behind after protein secretion and cleavage by signal peptidases. SppA is found in plants, archaea and bacteria. Here, we report the first crystal structure of a Gram-positive bacterial SppA. The 2.4-Å-resolution structure of *Bacillus subtilis* SppA ($SppA_{BS}$) catalytic domain reveals eight $SppA_{BS}$ molecules in the asymmetric unit, forming a dome-shaped octameric complex. The octameric state of $SppA_{BS}$ is supported by analytical size-exclusion chromatography and multi-angle light scattering analysis. Our sequence analysis, mutagenesis and activity assays are consistent with Ser147 serving as the nucleophile and Lys199 serving as the general base; however, they are located in different region of the protein, more than 29 Å apart. Only upon assembling the octamer do the serine and lysine come within close proximity, with neighboring protomers each providing one-half of the catalytic dyad, thus producing eight separate active sites within the complex, twice the number seen within *Escherichia coli* SppA ($SppA_{EC}$). The $SppA_{BS}$ S1 substrate specificity pocket is deep, narrow and hydrophobic, but with a polar bottom. The S3 pocket, which is constructed from two neighboring proteins, is shallower, wider and more polar than the S1 pocket. A comparison of these pockets to those seen in $SppA_{EC}$ reveals a significant difference in the size and shape of the S1 pocket, which we show is reflected in the repertoire of peptides the enzymes are capable of cleaving.

Introduction

Secretory proteins contain a signal peptide at their amino-terminus that facilitates transport to and translocation across the cytoplasmic membrane.[1,2] Once translocated, the signal peptide is cleaved off by signal peptidase. The remnant membrane-embedded signal peptide is subsequently degraded by signal peptide peptidase (Spp). Spp is a membrane-bound peptidase that is found in all cells. Although all Spp share the same function—digesting signal peptides—their size, membrane topology and catalytic mechanisms can be quite different. In archaea, plant chloroplasts and bacteria, these enzymes [signal peptide peptidase A (SppA)] utilize a serine nucleophile and a lysine general base.[3–7] However, in eukaryotic cells, these enzymes are aspartic proteases.[8] This report focuses on the Gram-positive bacterial serine protease *Bacillus subtilis* SppA ($SppA_{BS}$).

**Corresponding author*. E-mail address: mpaetzel@sfu.ca.

Abbreviations used: SppA, signal peptide peptidase A; SEC, size-exclusion chromatography; MALS, multi-angle light scattering; MPD, 2-methyl-2,4-pentanediol; CLS, Canadian Light Source.

Hussain *et al*. were the first to identify SppA as an enzyme involved in signal peptide digestion when they observed, in an *in vitro* experiment, that *Escherichia coli* lipoprotein signal peptides were digested upon the addition of a membrane extract containing SppA.[9] Bolhuis *et al*. were the first to report an SppA from *B. subtilis*. They showed that $SppA_{BS}$ can cleave the signal peptide of secretory proteins in an *in vivo* assay.[3]

Crystallographic analysis has revealed that the *E. coli* SppA ($SppA_{EC}$) contains two domains that are structurally a tandem repeat, although the sequence identity between the N- and C-terminal domains is only 18%.[7] The structure has also revealed that the enzyme forms a tetramer, with four separate active sites. Each active site being located at the interface between the N- and C-terminal domains, the catalytic serine arrives from the C-terminal domain, and the catalytic lysine arrives from the N-terminal domain of the same molecule. Mutagenesis experiments are consistent with $SppA_{EC}$ utilizing Ser409 as the nucleophile and Lys209 as the general base.[10] Based on topology predictions and analogy with $SppA_{EC}$, the catalytic domain of $SppA_{BS}$ is expected to be located on the trans-side of the cytoplasmic membrane with a single amino-terminal transmembrane segment anchoring it to the membrane.[10]

Here, we report the first X-ray crystal structure of a Gram-positive bacterial SppA. $SppA_{BS}$ assembles into an octameric dome-shaped structure, which creates eight separate active sites, as opposed to the tetrameric complex with four active sites observed in $SppA_{EC}$. The $SppA_{BS}$ active sites are constructed from the interface of neighboring protomers, suggesting that a proteolytically active enzyme is only produced upon assembly of an oligomeric $SppA_{BS}$. The S1 and S3 substrate specificity pockets (Schechter and Berger nomenclature[11]) are identified, described and compared to those in $SppA_{EC}$. We find that the dimensions and polarity of the S1 pocket differ significantly between $SppA_{BS}$ and $SppA_{EC}$ and that these differences are consistent with the range of residues the enzymes will accommodate at the P1 position of substrates.

Results and Discussion

$SppA_{BS}$ purification, crystallization and structure solution

We have observed that expressing the soluble domain of $SppA_{BS}$ ($SppA_{BS}\Delta 1–25$) in the cytoplasm of *E. coli* results in slow cell growth and a limited protein expression level. Sequence alignments (Supplementary Fig. 1) suggest that Ser147 serves as the nucleophile and that Lys199 served as the general base in the $SppA_{BS}$ catalytic mechanism; therefore, we designed the mutants K199A and S147A to test this hypothesis and to observe their effect on $SppA_{BS}$ expression level. Consistent with their proposed role as active-site residues, our activity assays reveal that these mutant enzymes are inactive. In addition, the K199A mutant was highly overexpressed in the cytoplasm of *E. coli* and was easily purified in milligram quantities. This protein was subjected to limited proteolysis using thermolysin in order to improve crystallization. The resulting proteolytically resistant fragment has a molecular mass of approximately 28 kDa, approximately 8 kDa smaller than the originally purified protein. Amino-terminal sequencing analysis has revealed that the proteolytically resistant fragment starts at Leu51 (Fig. 1a). The alcohols t-butanol and 2-methyl-2,4-pentanediol (MPD) and the detergent *n*-dodecyl-β-maltoside were used to produce conditions that led to highly ordered plate-shaped crystals. The 2.4-Å-resolution structure reveals clear electron density for residues 57–295, with the exception of a loop region between residues 72 and 82.

The $SppA_{BS}$ protein fold

The protomeric unit of the $SppA_{BS}$ catalytic domain (residues 57–295) has an α/β protein fold consisting of seven β-strands, eight α-helices and one 3_{10} helix (Fig. 1b). The $SppA_{BS}$ protomer has two regions: a globular region and an extension region. The globular region includes β-strands 1–4 and 7 that are arranged in a parallel β-sheet, surrounded by α-helices 1–3 and 6–8. The extension region consists of α-helices 4 and 5 and β-strands 5 and 6. Each protomer contains a nucleophilic residue, Ser147, and a general base residue, Lys199 (Fig. 1c). Ser147 is located at the turn before α-helix 3 of the globular region while Lys199 is located on the loop between β-strand 6 and α-helix 5 of the extension region (Fig. 1b and c). The distance from the Ser147 O^{γ} to the Lys199 C^{β} is approximately 29 Å, suggesting that a monomeric $SppA_{BS}$ would not be capable of catalysis. Eight molecules are in the asymmetric unit of this crystal (Fig. 2a). Analysis of the asymmetric unit reveals that the active-site catalytic dyad is created by Ser147 from one protomer and Lys199 from the neighboring protomer (Fig. 2b). Eight $SppA_{BS}$ protomers come together to form a dome-shaped structure with eight separate active sites.

$SppA_{BS}$ is octameric in solution

To confirm the existence of the octameric state of $SppA_{BS}$ in solution, we analyzed its size by analytical size-exclusion chromatography (SEC) and multi-angle light scattering (MALS). We observe that $SppA_{BS}$ has an average molecular mass of 225,400 ± 4,508 g/mol, consistent with the $SppA_{BS}$

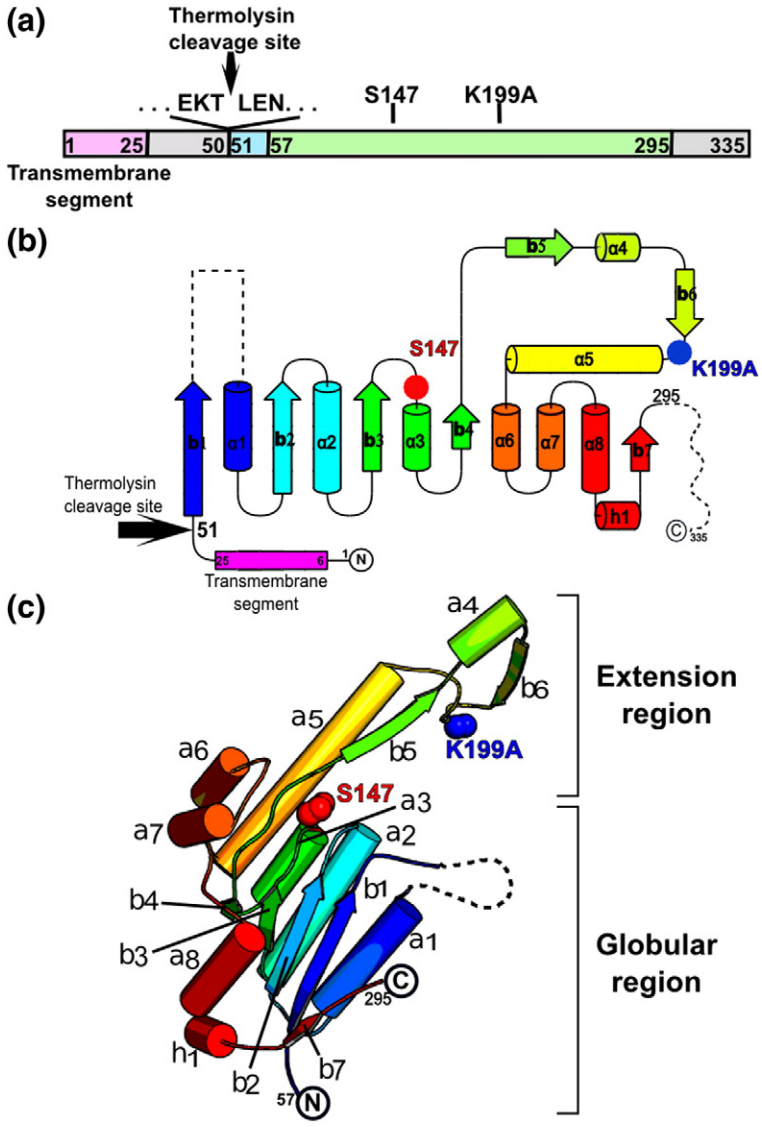

Fig. 1. The $SppA_{BS}$ protein folds. (a) A schematic diagram showing the full-length $SppA_{BS}$ with its predicted transmembrane segment (pink) and the confirmed amino-terminal thermolysin cleavage site. The light-green region is what is observed in the electron density. (b) A topology diagram of $SppA_{BS}$ showing the full-length protein with β-strand as arrows and α-helices as cylinders. The protein is colored as a gradient from N-terminus (blue) to C-terminus (red). The regions not seen in the electron density are shown as broken lines. (c) A cartoon diagram showing the tertiary structure of the $SppA_{BS}$ protomer. The β-strands are shown as arrows, and the α-helices are shown as cylinders. The color scheme is the same as in (b). The nucleophile Ser147 (red) and the general base Lys199Ala (blue) are shown as spheres.

complex (proteolytically resistant fragment) being in an octameric arrangement in solution (Fig. 2c).

The $SppA_{BS}$ dimensions and surface properties

The octameric $SppA_{BS}$ catalytic domain is dome shaped with a wide opening created by the globular regions and a narrower opening made by the extension regions (Fig. 3). Based on membrane topology predictions and analogy to $SppA_{EC}$, whose membrane topology has been experimentally determined, the wider opening of $SppA_{BS}$ is predicted to face the outer leaflet of the cytoplasmic membrane[10] (Fig. 3a). The outside of the dome has a relatively polar surface made up of both positively and negatively charged patches (Fig. 3b–d). The interior of the dome is predominantly hydrophobic with some patches of negative charge (Fig. 3c and e). The positively charged rim at the narrower opening of the dome is created by eight lysines, Lys185 from each of the $SppA_{BS}$ protomers (Fig. 3b). The narrow opening is ~22 Å in diameter. A cross-section of the dome shows a concave groove where the active sites and substrate specificity binding pockets reside. At this point, the diameter is approximately ~73 Å, while the narrowest region of the $SppA_{BS}$ dome interior is approximately 50 Å in diameter. The height of the dome is approximately 50 Å. The base of the dome, 86 Å in diameter at the opening, is primarily positively charged (Fig. 3e). The inner cavity has a solvent-accessible surface area of 17,220 $Å^2$ and a volume of 56,456 $Å^3$.

The $SppA_{BS}$ Ser/Lys catalytic dyad and other active-site residues

The catalytic dyad of $SppA_{BS}$ is made up of the nucleophilic Ser147 O^{γ} from one protomer and the

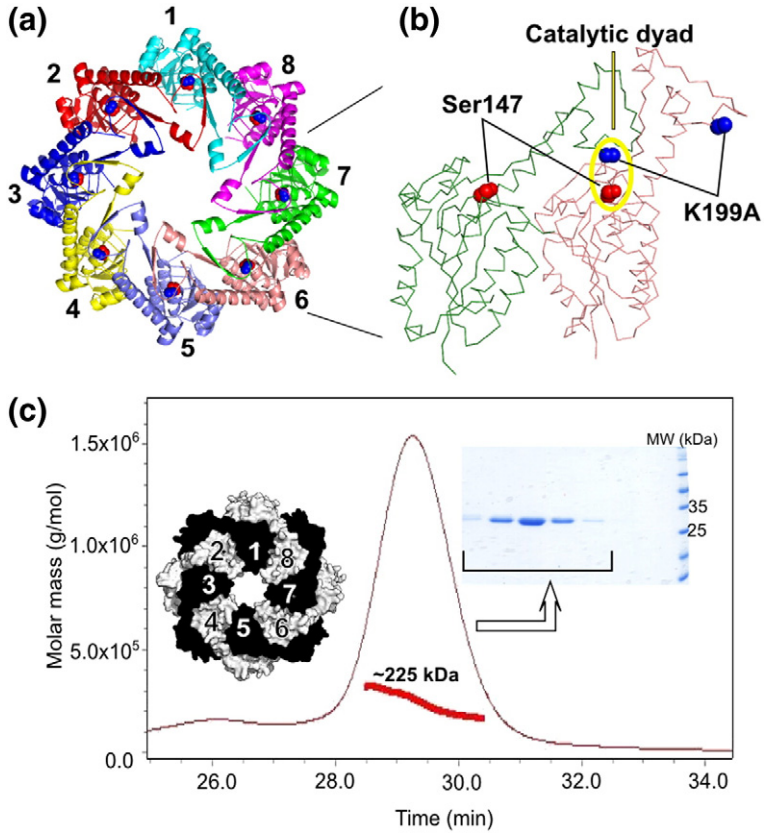

Fig. 2. Octameric arrangement of $SppA_{BS}$. (a) The $SppA_{BS}$ octamer as shown in a cartoon diagram, viewed from the top (closest to the extension region that forms the smaller opening). Each protomer is represented in a different color. The serine nucleophile (red) and lysine general base mutated to alanine (blue) are shown as spheres, revealing eight active-site catalytic dyads. (b) A C^{α} trace of two neighboring protomers coming together to form one of the eight active sites: the serine (red sphere) comes from one protomer, and the lysine (blue sphere) arrives from the adjacent protomer. (c) A MALS analysis of purified $SppA_{BS}$ proteolytic resistant fragment is represented by a molar mass (g/mol) *versus* time (min) plot overlaid with an analytical size-exclusion elution profile. A molecular mass of 225,000 Da is approximately equivalent to eight $SppA_{BS}$ protomers. The SDS-PAGE gel stained with PageBlue shows the fractions collected from the SEC column. The molecular mass markers are shown on the right. Next to the peak is a top view of the $SppA_{BS}$ octamer as shown in molecular surface representation, each protomer is colored alternating black and white.

general base Lys199 N^{ζ} from the adjacent protomer. Having eight protomers in the octamer therefore creates eight separate active sites, each formed at the protein–protein interface (Fig. 2). In order for $SppA_{BS}$ to have one completely functional unit, three neighboring $SppA_{BS}$ protomers are required to assemble (Fig. 4a). The nucleophile (Ser147), oxyanion hole residues (Gly114 and Gly148), the general base coordinating residue (Ser169) and S1 specificity pocket residues arrive from the central molecule (salmon in Fig. 4); the lysine general base arrived from one neighboring protomer (green in Fig. 4), and the S3 specificity pocket is partially formed from the other neighboring protomer (light blue in Fig. 4).

Modeling the Lys199 side chain within this K199A mutant structure shows that the Ser147 O^{γ} would be expected to be within hydrogen bonding distance to the Lys199 N^{ζ}. The oxyanion hole is created by the hydrogen bond donor NH groups of Gly114 and Gly148 from the same molecule that contributes the nucleophilic Ser147 to the catalytic dyad. The amino-terminal end of the α3 helix dipole, where Ser147 is located, is also a possible contributor to oxyanion stabilization. This molecule also contributes the side-chain hydroxyl group of Ser169, to the active site. Similarly positioned hydroxyl groups in other Ser/Lys proteases have been proposed to help coordinate and orient the lysine general base epsilon-amino group[12,13] (Fig. 4).

The $SppA_{BS}$ substrate specificity binding pockets

The eight active sites and substrate binding grooves of $SppA_{BS}$ are located approximately midway up the interior of the octameric dome creating a continuous concave surface that encircles the entire inner bowl (as seen in surface cross-section in Fig. 3e). Analysis of the grooves reveals only two clear specificity pockets for each of the eight active sites. Modeling an extended conformation for the substrate puts the P1 and P3 residue side chains facing the S1 and S3 specificity pockets and the P2 and P4 residue side chains facing toward the solvent away from the binding groove. The S1 substrate specificity pocket of $SppA_{BS}$ is created by residues Gly114, Val116, Ser119, Gly148, Tyr151, Gly171, Val172, Ser223, Gly226 and Phe227 from one protomer and Glu164 from a neighboring protomer while the S3 pocket is created by Val116, Met219 and Ser223 from one protomer and Pro163, Glu164, Thr165, Leu166, Asp252 and Arg254 from a third adjacent protomer (Fig. 4). The S1 pocket is narrow and deep with hydrophobic walls (Phe227, Tyr151 and Val116) and a more polar bottom (Glu164 $O^{\varepsilon 1}$,

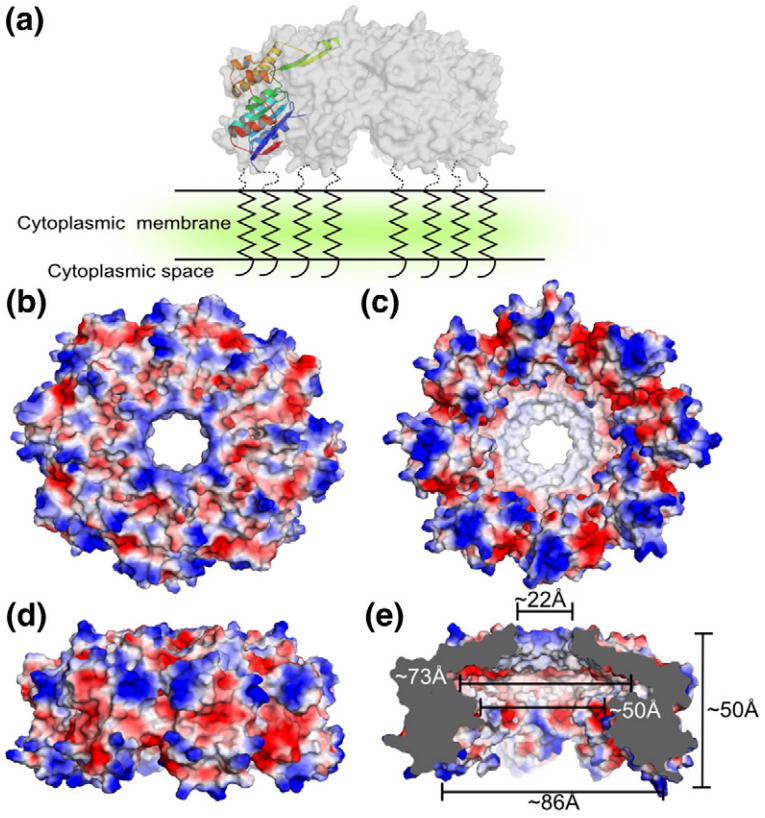

Fig. 3. Surface properties and dimensions of $SppA_{BS}$. (a) A semi-transparent surface representation of the $SppA_{BS}$ octamer, with one protomer shown in cartoon diagram. The widest opening in $SppA_{BS}$, formed from the globular region ("bottom"), is facing the cytoplasmic membrane, and the transmembrane segment is schematically drawn. The electrostatic properties of $SppA_{BS}$ are shown from the top (b), bottom (c) and side (d) (blue, positive; red, negative). (e) A cross-section view of the $SppA_{BS}$ octamer, with dimensions given.

$O^{\varepsilon 2}$ and Tyr151 O^{η}) (Fig. 4). The S3 pocket is not as deep as the S1 pocket but is significantly wider. The S3 entrance is hydrophobic however deeper into the pocket it has polar characteristics formed by the main-chain carbonyl oxygens of Pro163, Glu164 and Met219 and side chains of Ser223, Asp252 and Arg254. Structural alignment of the residues involved in creating the S1 and S3 pockets in the eight binding sites of $SppA_{BS}$ showed that they superimpose quite well except for residues Ser223, Glu164 and Arg254 (Fig. 4c). The Ser223 side-chain rotamer, which is located at the S1 and S3 pocket boundary, varies among the eight chains. Alternate conformations are observed for the side chains of residues Glu164 and Arg254, which form part of the S3 pocket, this suggests that these residues maybe dynamic and may possibly be involved in an induced fit process upon substrate binding.

Octameric $SppA_{BS}$ *versus* tetrameric $SppA_{EC}$

Both $SppA_{BS}$ and $SppA_{EC}$ form oligomers that assemble into dome-shaped structures.[7] The $SppA_{BS}$ protomer is half the size of the $SppA_{EC}$ protomer. Therefore, the $SppA_{BS}$ dome complex is created by eight protomers while only four protomers are required for $SppA_{EC}$ (Fig. 5a and b). $SppA_{EC}$'s N- and C-terminal domains are structurally tandem repeats (Fig. 5c). The $SppA_{BS}$ protomer superimposes onto the N-terminal domain of $SppA_{EC}$ with an r.m.s.d. value of 2.7 Å (Fig. 5e) while the $SppA_{BS}$ protomer superimposes onto the C-terminal domain of $SppA_{EC}$ with a much lower r.m.s.d. values of 1.1 Å (Fig. 5f). The most significant differences in the superimposition of the N-terminal domain of $SppA_{EC}$ with the $SppA_{BS}$ protomer are observed in the outer helices (α-helices 1 and 8) and the extension region. In addition, there is an extra helix found in the N-terminal domain of $SppA_{EC}$, between α-helices 6 and 7 of $SppA_{BS}$ (Fig. 5e).

Sequence conservation between Gram-positive and Gram-negative SppA

$SppA_{BS}$ has a sequence identity of 16% and 26% to the N- and C-terminal domains of $SppA_{EC}$, respectively (Supplementary Fig. 1). The majority of the conserved residues shared by $SppA_{BS}$ and the C-terminal domain of $SppA_{EC}$ are evenly distributed on the protomer of $SppA_{BS}$; however, there is a patch of conserved residues clustered around the nucleophile Ser147 (Supplementary Fig. 2). Similar patterns of conserved residues are also observed when $SppA_{BS}$ is aligned with SppA sequences from other Gram-positive species (Supplementary Figs. 3 and 5) or when $SppA_{EC}$ is aligned with SppA sequences from other Gram-negative species (Supplementary Figs. 4 and 5).

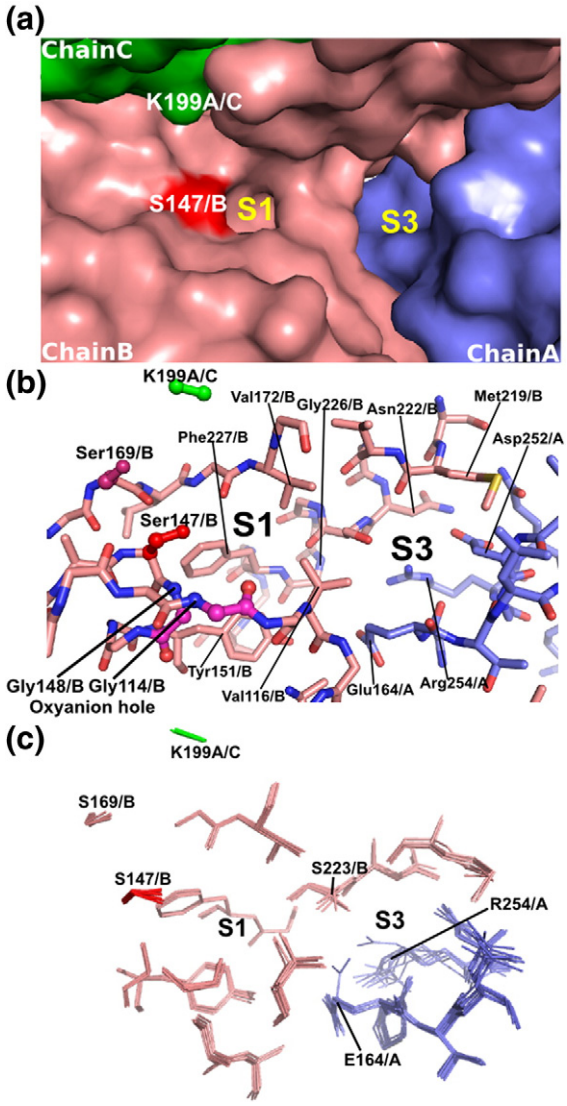

Fig. 4. $SppA_{BS}$ catalytic residues and substrate specificity pockets. (a) Three protomers of $SppA_{BS}$ create one complete catalytic core with the general base (K199A; green) arriving from the green-colored protomer, the nucleophile (Ser147; red) and residues that make up the S1 substrate specificity pocket arriving from the salmon-colored protomer and the S3 pocket being formed from residues in the salmon-colored and light-blue-colored protomers. (b) Stick representation of residues involved in forming the active-site and substrate specificity pockets. Ser147 is shown in red, K199A is shown in green and the proposed general base orienting residue Ser169 is shown in maroon. Gly114 and Gly148 shown in magenta form the oxyanion hole. (c) Superimposition of residues involved in forming each of the eight active sites and binding pockets.

SppA of Gram-positive bacteria are located on the extracellular surface, whereas SppA of Gram-negative bacteria are located on the inner membrane facing the periplasmic space. Given the potentially great surface exposure of Gram-positive SppA, one might expect there to be less conservation on their exposed surfaces, as compared to the Gram-negative SppA. Mapping residue conservation onto the surface of the $SppA_{BS}$ and $SppA_{EC}$ structures, based on separate Gram-positive and Gram-negative SppA sequence alignments (Supplementary Figs. 3 and 4), reveals a distinctly different pattern of conservation that is consistent with the potential difference in surface exposure. There are four patches of conservation that show up on the top surface of Gram-negative SppA, whereas a broader, more evenly distributed, circle of conservation is observed on the top surface of Gram-positive SppA (Fig. 6).

S1 substrate specificity binding pocket: $SppA_{BS}$ *versus* $SppA_{EC}$

A comparison of the substrate binding grooves of $SppA_{EC}$ and $SppA_{BS}$ reveals that the $SppA_{BS}$ S1 substrate specificity pocket is significantly deeper (Fig. 7). Structural alignment of the $SppA_{BS}$ and $SppA_{EC}$ active-site residues shows that the smaller S1 pocket seen in $SppA_{EC}$ results mostly from three residue substitutions: $SppA_{BS}$ Tyr151 to $SppA_{EC}$ Trp413, $SppA_{BS}$ Ala120 to $SppA_{EC}$ Glu383 and $SppA_{BS}$ Gly226 to $SppA_{EC}$ Arg487 (Fig. 7c).

Range of residues accommodated at the substrate P1 position: $SppA_{BS}$ *versus* $SppA_{EC}$

Using wild-type active-site enzymes and a series of peptide-MCA (4-*m*ethyl 7-*c*umaryl *a*mide) fluorogenic substrates (where the C-terminal residue corresponds to the P1 residue), we observe that both $SppA_{EC}$ and $SppA_{BS}$ show a preference for the leucine tripeptide (LLL) substrate. This is consistent with the predominance of leucine residues in the H-region of signal peptides. Yet a clear difference can be observed between $SppA_{EC}$ and $SppA_{BS}$ when using the other peptide-MCA substrates (Fig. 8). The second and third most effective substrates for $SppA_{BS}$ in this series are LRR and LLVY, respectively, whereas $SppA_{EC}$ shows close to no detectible activity using the same compounds. The ability of $SppA_{BS}$ to cleave the LRR and LLVY peptides, where arginine and tyrosine are the P1 residues, is consistent with its deeper and more polar S1 pocket (Fig. 7). No detectable activity was observed for the $SppA_{BS}$ active-site mutant K199A or for $SppA_{EC}$ with the corresponding active-site mutation (K209A).

The $SppA_{EC}$ and $SppA_{BS}$ S1 pocket shape differences along with the corresponding observed differences in the range of acceptable P1 residues are somewhat surprising given that the signal peptides (both predicted and experimentally verified) from *B. subtilis* and *E. coli* reveal approximately the same residue content, the only significant

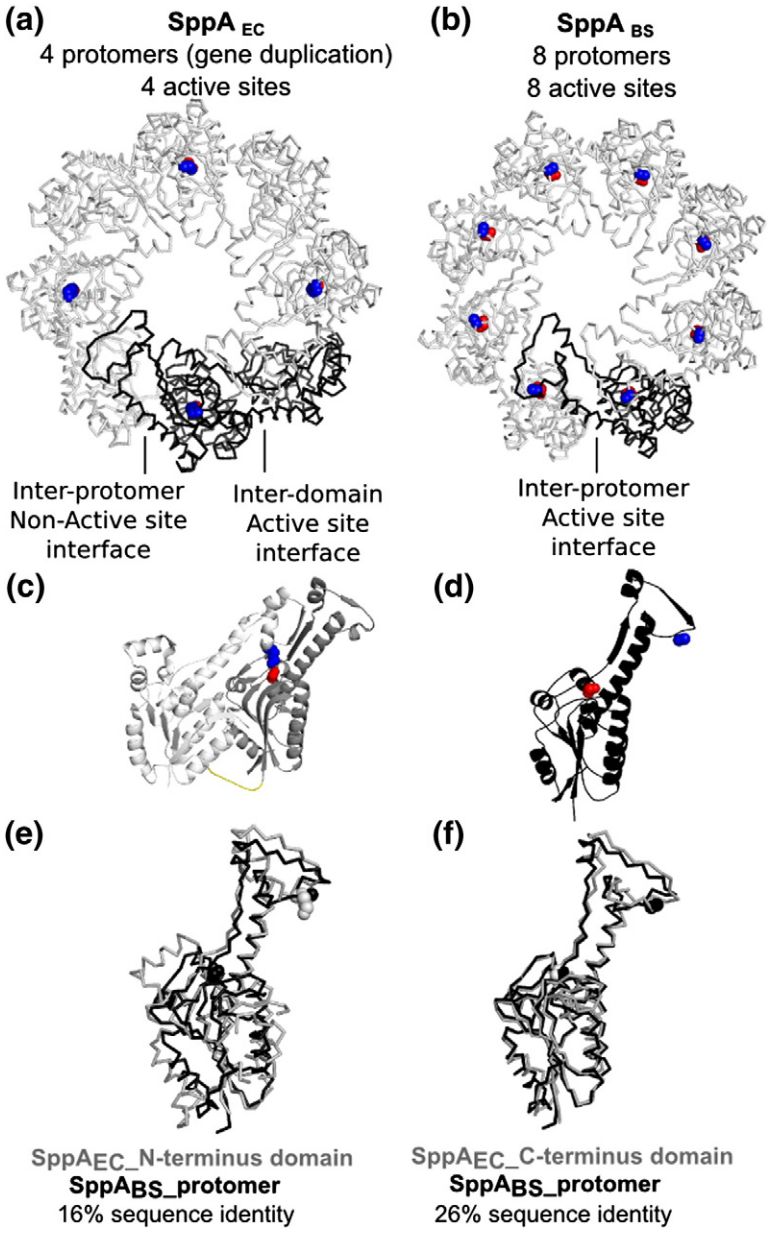

Fig. 5. Comparison between octameric $SppA_{BS}$ and tetrameric $SppA_{EC}$. (a) The C^{α} trace of $SppA_{EC}$. (b) The C^{α} trace of $SppA_{BS}$. The catalytic dyads are shown (general base, blue spheres; nucleophile, red spheres). One protomer in each oligomer is shown in black in (a) and (b). (c) A cartoon diagrams of an $SppA_{EC}$ protomer: the C-terminal domain is colored darker gray, and the region linking the domains is colored yellow. (d) A cartoon diagrams of an $SppA_{BS}$ protomer. (e) A superposition of the $SppA_{BS}$ protomer (black) on the N-terminal domains of $SppA_{EC}$ (gray). (f) A superposition of the $SppA_{BS}$ protomer (black) on the C-terminal domains of $SppA_{EC}$ (gray).

difference appears to be that the signal peptides of Gram-positive bacteria are in general longer than those in Gram-negative bacteria.[14] Given the differences between the substrate specificity pockets, it is plausible that SppA may have other substrates in addition to signal peptides.

SppA is capable of digesting folded proteins

Interestingly, we observed that the soluble domains of both $SppA_{BS}$ and $SppA_{EC}$ are capable of digesting fully folded proteins. For example, both enzymes digest the *E. coli* lipoprotein BamD, a periplasmic protein involved in outer membrane protein assembly, to near completion after a 17-h incubation. The active-site mutant forms of each enzyme showed no ability to digest BamD (Supplementary Fig. 6). The ability of SppA to digest folded proteins and to cleave peptides is consistent with SppA possibly having a membrane protein quality assurance role. Experiments are currently underway to explore this possibility.

Experimental Procedures

Cloning and mutagenesis

The *B. subtilis sppA* gene, lacking the nucleotides that code for residues 1–25 ($SppA_{BS}\Delta 1–25$), was amplified

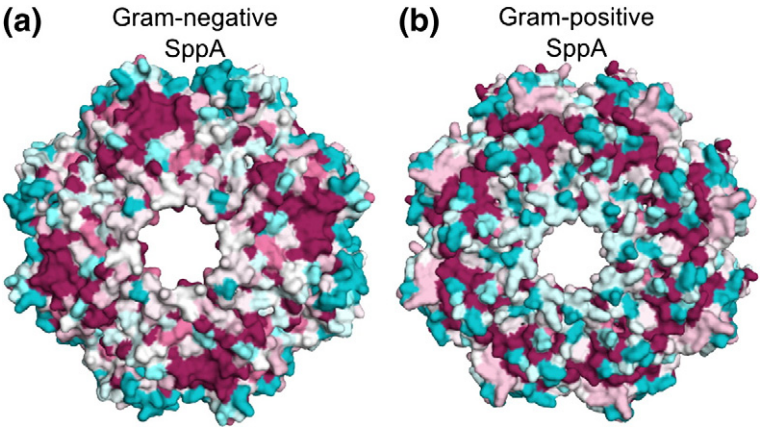

Fig. 6. Surface residue conservation of Gram-negative SppA *versus* Gram-positive SppA. Residue conservation of Gram-negative SppA and Gram-positive SppA are mapped onto the surface of the $SppA_{EC}$ and $SppA_{BS}$ structures, as viewed from the top external surface. The conservation is based on separate Gram-positive and Gram negative SppA sequence alignments (Supplementary Figs. 3 and 4). Completely conserved residues are shown in maroon, while highly variable residues are shown in cyan.

using PCR. The oligonucleotides used to amplify this construct were forward primer 5′ gccatatgagtttctttgaaagcgtcaaaggc 3′ and reverse primer 5′ ctcgagctacttcgcatagagatacatcattct 3′. The $SppA_{BS}\Delta 1–25$ construct was cloned into the expression vector pET28b+ (Novagen) using the NdeI and XhoI restriction sites. Using the above construct as a template, we individually mutated the codons for the proposed active-site residues Lys199 and Ser147 to a codon for alanine following the QuikChange site-directed mutagenesis procedure. The oligonucleotides used to perform the K199A site-directed mutagenesis were forward primer 5′ agcggggcccatgcggacattatgtct 3′ and reverse primer 5′ agacataatgtccgcatgggccccgct 3′. The oligonucleotides used to perform the S147A site-directed mutagenesis were forward primer 5′ tcgatggcagcagcaggaggctattac 3′ and reverse primer 5′ gtaatagcctcctgctgctgccatcga 3′. The sequences of the native active site, the K199A mutant and the S147A mutant constructs were confirmed by DNA sequencing. The expressed proteins have a hexa-histidine affinity tag followed by a thrombin cleavage site such that the amino-terminus of the expressed protein has the following sequence preceding the 26th residue of the *B. subtilis sppA* gene (MGSSHHHHHHSSGLVPRGSH). The expressed K199A mutant protein (including 6× His tag and linker/thrombin site) has a calculated molecular mass of 36,182 Da and a theoretical isoelectric point (p*I*) of 6.9.

In addition, another construct of the *B. subtilis sppA* gene was prepared, and this construct lacks the nucleotides that code for residues 2–54 ($SppA_{BS}\Delta 2–54$). This wild-type active-site construct gave a sufficient expression level for the enzyme to be purified and for activity assays to be performed. The construct was amplified using PCR. The oligonucleotides used to amplify this construct were forward primer 5′ catatgagtccctcaagtaaaattgccg 3′ and reverse primer 5′ ctcgagctacttcgcatagagatacatcattct 3′. The $SppA_{BS}\Delta 2–54$ construct was cloned into the expression vector pET28a+ (Novagen) using the NdeI and XhoI restriction sites. Active-site mutations (S147A and K199A) in this construct were prepared as described above for the $SppA_{BS}\Delta 1–25$ construct. The sequences were confirmed by DNA sequencing. The expressed proteins have a hexa-histidine affinity tag followed by a thrombin cleavage site such that the amino-terminus of the expressed protein has the following sequence preceding the 55th residue of the *B. subtilis sppA* gene (MGSSHHHHHHSSGLVPRGSH). The expressed native active-site protein (including 6× His tag

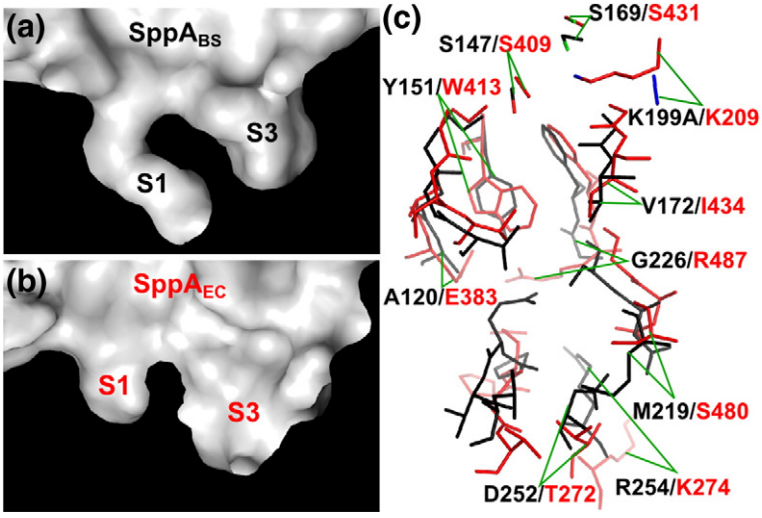

Fig. 7. Substrate specificity pocket comparison between $SppA_{BS}$ and $SppA_{EC}$. (a) A cross-section of the substrate specificity groove surface in $SppA_{BS}$. (b) A cross-section of the substrate specificity groove surface in $SppA_{EC}$. (c) A superimposition of the binding-site residues in $SppA_{BS}$ (black) and $SppA_{EC}$ (red), rendered in stick.

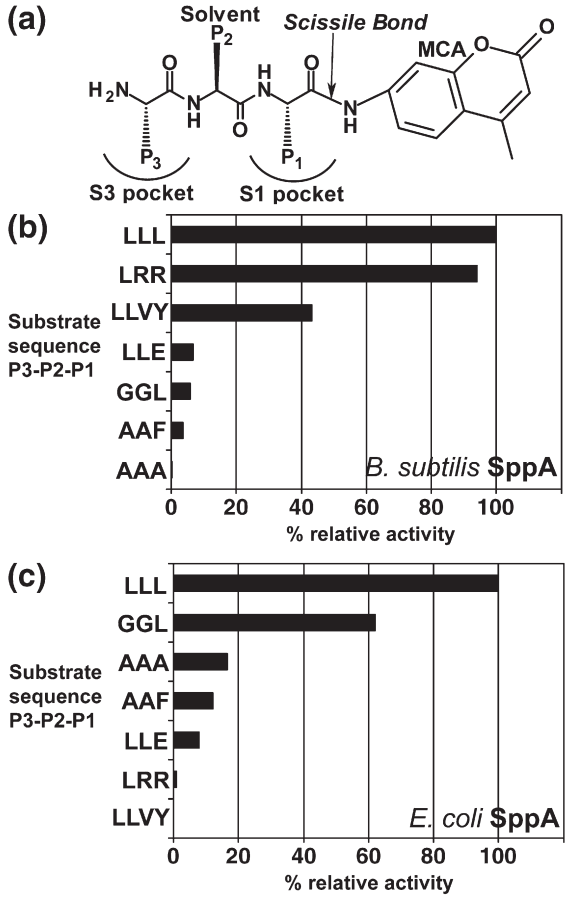

Fig. 8. Activity profile of $SppA_{BS}$ *versus* $SppA_{EC}$ reveals a difference in the range of residues the enzymes will accommodate at the P1 position. (a) A schematic diagram of the fluorogenic peptide-MCA substrates used to compare % relative activity in $SppA_{BS}$ and $SppA_{EC}$. (b) The activity profile for $SppA_{BS}$. (c) The activity profile for $SppA_{EC}$.

and linker/thrombin site) has a calculated molecular mass of 33,137 Da and a theoretical isoelectric point (p*I*) of 8.7.

Expression and purification

E. coli Tuner (DE3) competent cells were transformed with the plasmids described above. The cells were grown in LB media containing kanamycin (0.05 mg/ml) at 37 °C to an OD_{600} of 0.6 and induced with 0.5 mM IPTG at 25 °C for 1 h. The cells were harvested by centrifugation (6,000*g* for 6 min) and then resuspended in buffer A [20 mM Tris–HCl (pH 8.0) and 150 mM NaCl]. The cells were lysed by sonication with three 15-s pulses at 30% amplitude with a 30-s rest between each of the pulses (Fisher Scientific Sonic Dismembrator Model 500). Homogenization was then carried out at 20,000–25,000 psi for 3 min using an Avestin Emulsiflex-3C cell homogenizer. The cell lysate was centrifuged at 30,000*g* for 35 min, and the supernatant was applied to a Ni^{2+}–NTA affinity chromatography column pre-equilibrated with buffer A. The column was washed with 50 mM and 75 mM imidazole in buffer A. The protein was eluted using a step-wise gradient, 100–600 mM imidazole in buffer A. Fractions containing $SppA_{BS}$ were pooled and dialyzed against buffer A overnight at 4 °C to remove imidazole. Dialyzed protein was then concentrated to 10 mg/ml using an Amicon ultracentrifugal filter device (Millipore) with a 10-kDa cutoff. The protein concentration was measured using a Nanodrop ND-100 spectrophotometer. The extinction coefficient (17,880 M^{-1} cm^{-1} for $SppA_{BS}\Delta1$–25_K199A) was calculated based on the $SppA_{BS}$ amino acid sequence using *ProtParam*.[15]

Limited proteolysis

$SppA_{BS}\Delta1$–25_K199A (10 mg/ml) was digested with thermolysin (Sigma) (500:1 molar ratio) overnight at room temperature. The sample was then applied to a Superdex 200 SEC column, equilibrated with buffer A, on an Amersham ÄKTA FPLC system running at a flow rate of 0.5 ml/min. The fractions containing the stable proteolytic fragment of $SppA_{BS}$ were pooled and concentrated to ~18 mg/ml.

Amino-terminal sequencing analysis

To determine the amino-terminus created by thermolysin digestion, we dissolved crystals of the thermolysin digested $SppA_{BS}$ in buffer A, run on 13.5% SDS-PAGE and blotted electrophoretically onto a PDVF membrane (Millipore). The membrane was stained with Coomassie Brilliant Blue R-250 and sent to the Iowa State University Protein Facility. Automated Edman sequencing with a model 494 Procise Protein Sequencer/140C Analyzer (Applied Biosystems, Inc.) was used to determine the sequence for the first six amino acid residues.

Analytical SEC and MALS analysis

To determine the oligomeric state of $SppA_{BS}$ in solution, we applied thermolysin treated $SppA_{BS}$ (10 mg/ml) to a Superdex 200 (GE Healthcare) size-exclusion column equilibrated with buffer A and run at a flow rate of 0.4 ml/min. The column was connected inline to a Dawn 18-angle light-scattering detector coupled to an Optilab rEX interferometric refractometer and a quasi-elastic light-scattering instrument (Wyatt Technologies). The molecular mass was calculated with ASTRA v5.1 software (Wyatt Technologies, Inc.) using the Zimm fit method[16] with a refractive index increment, $dn/dc = 0.185$ ml/g.

SppA activity assay via fluorogenic peptide substrates

The reactions were carried out in buffer A at 37 °C using a Cary Eclipse fluorescence spectrophotometer (Varian). A series of peptide-MCA fluorogenic substrates were prepared at 10 mM stock solution by dissolving them in 100% dimethyl sulfoxide. The substrates Z-AAF-MCA, Suc-AAA-MCA, Boc-LRR-MCA and Z-LLL-MCA were

purchased from the Peptide Institute. The substrates Suc-LLVY-MCA and Z-LLE-MCA were purchased from Calbiochem. The substrate Z-GGL-MCA was purchased from Bachem. The abbreviations for the amino-terminal modifications are as follows: Z- is benzyloxycarbonyl, Suc- is succinyl and Boc- is t-butyloxycarbonyl. For comparison of the relative activity between $SppA_{BS}\Delta2–54$ and $SppA_{EC}\Delta2–46$, a single substrate concentration was utilized (10 μM for $SppA_{BS}\Delta2–54$ and 2 μM for $SppA_{EC}\Delta2–46$) in each 500-μl reaction containing 40 nM enzyme in a 1 cm × 1 cm cuvette. The final dimethyl sulfoxide concentration was 2%. The excitation and emission wavelengths used were 380 nm and 460 nm, respectively.

Crystallization

The thermolysin digested $SppA_{BS}$ crystals were grown using the sitting-drop vapor diffusion method. The detergent *n*-dodecyl-β-maltoside (0.1% final concentration) was added to the protein before setting up crystallization drops. The drop contained 1 μl of protein and 1 μl of reservoir solution. The refined reservoir condition was 23% t-butanol, 0.1 M Tris–HCl (pH 8.5) and 5% MPD. The drop was equilibrated against 1 ml of reservoir solution at 18 °C. The cryo-solution contained 20% MPD, 23% t-butanol and 0.1 M Tris–HCl (pH 8.5). The crystal was transferred to the cryo-solution and flash-cooled in liquid nitrogen.

Diffraction data collection

Diffraction images were collected on beamline 08B1-1 at the Canadian Macromolecular Crystallography Facility of the Canadian Light Source (CLS), using a Rayonix MX300HE X-ray detector. A total of 360 images were collected at wavelength 0.9795 Å with 0.5° oscillations, and each image was exposed for 1 s. The crystal-to-detector distance was 250 mm. HKL2000 was used to process the diffraction images.[17] The crystal belongs to space group $P2_12_12_1$ with unit cell dimensions of 87.8 Å × 131.1 Å × 207.3 Å. There are eight molecules in the asymmetric unit with a Matthews coefficient of 2.67 $Å^3$ Da^{-1} (54.0% solvent), which was calculated with the Matthews Probabilities calculator[18] using the $SppA_{BS}$ molecular mass after limited proteolysis (28,000 Da). See Table 1 for crystal parameters, data collection and refinement statistics.

Structure determination and refinement

Phase estimates were obtained by molecular replacement using the program *Phaser*.[19] A search model was built using the C-terminal domain of $SppA_{EC}$ (Protein Data Bank ID: 3BF0; chain A) as a template. The homology model was built using the program *CHAINSAW*.[20] The side chains of conserved residues were included, and non-conserved side chains were truncated to the C^{β} atom. The initial R_{work}/R_{free} before refinement was 0.3484/0.4696. The program Autobuild within *PHENIX* version 1.6.4[21] was used to build the side chains and refine the initial structure. Restrained refinement was performed using the program *REFMAC5*,[22] and manual adjustments and manipulations were executed with the program *Coot*.[23] The final round of refinement included TLS and restrained refinement with five TLS groups for each chain. Input files were obtained by the TLS motion determination server.[24,25] See Table 1 for crystal parameters, data collection statistics and refinement statistics.

Table 1. Data collection and refinement statistics

Crystal parameters	
Space group	$P2_12_12_1$
a, *b*, *c* (Å)	87.8, 131.1, 207.3
Data collection statistics	
Wavelength (Å)	0.9795
Resolution (Å)	50.0–2.4 (2.5–2.4)[a]
Total reflections	709,425
Unique reflections	96,535 (9520)
R_{merge}[b]	0.085 (0.294)
Mean $(I)/\sigma(I)$	52.7 (7.8)
Completeness (%)	99.8 (99.8)
Redundancy	7.4 (6.9)
Refinement statistics	
Protein molecules (chains) in asymmetric unit	8
Residues	1803
Water molecules	257
Total number of atoms	13,957
R_{cryst}[c]/R_{free}[d] (%)	20.6/24.1
Average *B*-factor ($Å^2$) (all atoms)	49.0
r.m.s.d. on angles (°)	1.135
r.m.s.d. on bonds (Å)	0.010

[a] The data collection statistics in parentheses are the values for the highest-resolution shell.

[b] $R_{merge} = \sum_{hkl}\sum_j |I_j(hkl) - (I(hkl))| / \sum_{hkl}\sum_j I_j(hkl)$, where $I_j(hkl)$ is the intensity of an individual reflection and is the mean intensity of that reflection.

[c] $R_{cryst} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

[d] R_{free} is calculated using 5% of the reflections randomly excluded from refinement.

Structural analysis

The figures were created using *PyMOL*.[26] *CLUSTALW*[27] and *ESPript* v.2.2[28] were utilized for the sequence alignments. Signal peptide sequences of *B. subtilis* and *E. coli* were obtained from SPdb Signal Peptide Resource.[29] The protein–protein interface between protomers was analyzed using *PISA*[30] and *ProtorP*.[31] The conserved residues among Gram-negative and Gram-positive SppA were mapped onto the structure using *ConSurf*.[32–34] The substrate specificity pockets were analyzed using *CASTp*[35] while the stereochemistry of the structure was analyzed with the program *PROCHECK*.[36] *PROMOTIF*[37] was used to identify and analyze the secondary structure and motifs within in the protein.

Accession numbers

Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 3RST.

Acknowledgements

This work was supported in part by the Canadian Institute of Health Research (to M.P.), the Natural Science and Engineering Research Council of Canada (to M.P.), the Michael Smith Foundation for Health Research (to M.P.) and the Canadian Foundation of Innovation (to M.P.). We thank the staff at beamline 08ID-1 at the CLS, Saskatoon, Canada (especially Shaunivan Labiuk and Julien Cotelesage) for their help with data collection. CLS was supported by Natural Science and Engineering Research Council of Canada, National Research Council, Canadian Institute of Health Research and the University of Saskatchewan. Special thanks to Dr. Jaeyong Lee for helpful discussions.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2012.03.020

References

1. Paetzel, M., Karla, A., Strynadka, N. C. & Dalbey, R. E. (2002). Signal peptidases. *Chem. Rev.* **102**, 4549–4580.
2. Yuan, J., Zweers, J. C., van Dijl, J. M. & Dalbey, R. E. (2010). Protein transport across and into cell membranes in bacteria and archaea. *Cell. Mol. Life Sci.* **67**, 179–199.
3. Bolhuis, A., Matzen, A., Hyyrylainen, H. L., Kontinen, V. P., Meima, R., Chapuis, J. *et al.* (1999). Signal peptide peptidase- and ClpP-like proteins of *Bacillus subtilis* required for efficient translocation and processing of secretory proteins. *J. Biol. Chem.* **274**, 24585–24592.
4. Lensch, M., Herrmann, R. G. & Sokolenko, A. (2001). Identification and characterization of SppA, a novel light-inducible chloroplast protease complex associated with thylakoid membranes. *J. Biol. Chem.* **276**, 33645–33651.
5. Matsumi, R., Atomi, H. & Imanaka, T. (2006). Identification of the amino acid residues essential for proteolytic activity in an archaeal signal peptide peptidase. *J. Biol. Chem.* **281**, 10533–10539.
6. Novak, P. & Dev, I. K. (1988). Degradation of a signal peptide by protease IV and oligopeptidase A. *J. Bacteriol.* **170**, 5067–5075.
7. Kim, A. C., Oliver, D. C. & Paetzel, M. (2008). Crystal structure of a bacterial signal peptide peptidase. *J. Mol. Biol.* **376**, 352–366.
8. Weihofen, A., Binns, K., Lemberg, M. K., Ashman, K. & Martoglio, B. (2002). Identification of signal peptide peptidase, a presenilin-type aspartic protease. *Science*, **296**, 2215–2218.
9. Hussain, M., Ozawa, Y., Ichihara, S. & Mizushima, S. (1982). Signal peptide digestion in *Escherichia coli*. Effect of protease inhibitors on hydrolysis of the cleaved signal peptide of the major outer-membrane lipoprotein. *Eur. J. Biochem.* **129**, 233–239.
10. Wang, P., Shim, E., Cravatt, B., Jacobsen, R., Schoeniger, J., Kim, A. C. *et al.* (2008). *Escherichia coli* signal peptide peptidase A is a serine–lysine protease with a lysine recruited to the nonconserved amino-terminal domain in the S49 protease family. *Biochemistry*, **47**, 6361–6369.
11. Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157–162.
12. Paetzel, M., Dalbey, R. E. & Strynadka, N. C. (2002). Crystal structure of a bacterial signal peptidase apoenzyme: implications for signal peptide binding and the Ser–Lys dyad mechanism. *J. Biol. Chem.* **277**, 9512–9519.
13. Feldman, A. R., Lee, J., Delmas, B. & Paetzel, M. (2006). Crystal structure of a novel viral protease with a serine/lysine catalytic dyad mechanism. *J. Mol. Biol.* **358**, 1378–1389.
14. van Roosmalen, M. L., Geukens, N., Jongbloed, J. D., Tjalsma, H., Dubois, J. Y., Bron, S. *et al.* (2004). Type I signal peptidases of Gram-positive bacteria. *Biochim. Biophys. Acta*, **1694**, 279–297.
15. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook* (Walker, J. M., ed.), pp. 571–607, Humana Press, Totowa, NJ.
16. Zimm, B. H. (1948). The scattering of light and the radial distribution function of high polymer solutions. *J. Chem. Phys.* **16**, 1093–1099.
17. Otwinowski, Z. (1993). Denzo. In *Denzo* (Sawyer, L., Isaacs, N. & Baily, S., eds), pp. 56–62, Daresbury Laboratory, Warrington, UK.
18. Kantardjieff, K. A. & Rupp, B. (2003). Matthews coefficient probabilities: improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Sci.* **12**, 1865–1871.
19. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). Likelihood-enhanced fast translation functions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **61**, 458–464.
20. Stein, N. (2008). CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr.* **41**, 641–643.
21. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N. *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 213–221.
22. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A. *et al.* (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **67**, 355–367.
23. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
24. Painter, J. & Merritt, E. A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **62**, 439–450.
25. Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374**, 300–321.

26. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA.
27. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
28. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: multiple sequence alignments in PostScript. *Bioinformatics*, **14**, 305–308.
29. Choo, K. H., Tan, T. W. & Ranganathan, S. (2005). SPdb—a signal peptide database. *BMC Bioinformatics*, **6**, 249.
30. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
31. Reynolds, C., Damerell, D. & Jones, S. (2009). ProtorP: a protein–protein interaction analysis server. *Bioinformatics*, **25**, 413–414.
32. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529–W533.
33. Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, **19**, 163–164.
34. Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. & Ben-Tal, N. (2005). ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–W302.
35. Liang, J., Edelsbrunner, H. & Woodward, C. (1998). Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* **7**, 1884–1897.
36. Laskowski, R. A. (2001). PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res.* **29**, 221–222.
37. Hutchinson, E. G. & Thornton, J. M. (1996). PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212–220.