doi:10.1016/j.jmb.2007.11.080 J. Mol. Biol. (2008) 376, 352–366

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Crystal Structure of a Bacterial Signal Peptide Peptidase

Apollos C. Kim, David C. Oliver 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 21 September 2007; received in revised form 20 November 2007; accepted 22 November 2007
Available online
4 December 2007

Signal peptide peptidase (Spp) is the enzyme responsible for cleaving the remnant signal peptides left behind in the membrane following Sec-dependent protein secretion. Spp activity appears to be present in all cell types, eukaryotic, prokaryotic and archaeal. Here we report the first structure of a signal peptide peptidase, that of the *Escherichia coli* SppA ($SppA_{EC}$). $SppA_{EC}$ forms a tetrameric assembly with a novel bowl-shaped architecture. The bowl has a dramatically hydrophobic interior and contains four separate active sites that utilize a Ser/Lys catalytic dyad mechanism. Our structural analysis of SppA reveals that while in many Gram-negative bacteria as well as characterized plant variants, a tandem duplication in the protein fold creates an intact active site at the interface between the repeated domains, other species, particularly Gram-positive and archaeal organisms, encode half-size, unduplicated SppA variants that could form similar oligomers to their duplicated counterparts, but using an octamer arrangement and with the catalytic residues provided by neighboring monomers. The structure reveals a similarity in the protein fold between the domains in the periplasmic Ser/Lys protease SppA and the monomers seen in the cytoplasmic Ser/His/Asp protease ClpP. We propose that SppA may, in addition to its role in signal peptide hydrolysis, have a role in the quality assurance of periplasmic and membrane-bound proteins, similar to the role that ClpP plays for cytoplasmic proteins.

Keywords: Signal peptide peptidase; protease IV; Ser/Lys dyad; Lysine general-base; serine protease

Edited by R. Huber

Introduction

The cellular strategy of using an amino-terminal extension (signal peptide) to help facilitate targeting and translocation of proteins across cellular membranes is a process conserved across all forms of life.[1] In order for secreted proteins to be released from the membrane after translocation the signal peptide is cleaved off by signal peptidase. It is the role of signal peptide peptidase to further cleave the signal peptides into smaller peptides that are able to be released from the membrane. Novak and Dev confirmed, based on their *in vitro* biochemical analysis of *Escherichia coli* signal peptide degradation, that *E. coli* signal peptide peptidase ($SppA_{EC}$) is the endopeptidase that cleaves within the hydrophobic region (H-region) of the signal peptide.[2]

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

Abbreviations used: Spp, signal peptide peptidase; $SppA_{EC}$, *Escherichia coli* SppA; SAD, single-wavelength anomalous diffraction.

$SppA_{EC}$, also referred to as protease IV, is 618 residues in length (molecular mass 67,219 Da).[3] Biochemical analysis has revealed that SppA is anchored to the cytoplasmic membrane.[4]

SppA homologues exist in eubacteria,[4,5] archaea[6] and the thylakoid membrane of chloroplasts[7] and they collectively belong to the evolutionary clan of proteases SK and the protease family S49.[8] Unlike the eukaryotic functional homologue that has been proposed to be an aspartic protease,[9] site-directed mutagenesis along with sequence analysis is consistent with the bacterial, archaeal and thylakoid SppA being a serine protease.[6]

Results and Discussion

A catalytically active and soluble version of the membrane protein $SppA_{EC}$ ($SppA_{EC}\Delta 2–46$) that lacks residues 2 through 46 (containing the predicted N-terminal transmembrane segment, residues 29–45) was used for crystallization. $SppA_{EC}\Delta 2–46$ was crystallized in the space group $P2_1$ with four

molecules in the asymmetric unit and its structure was solved by single-wavelength anomalous diffraction (SAD). The final refined structure, at 2.6 Å resolution, includes residues 56 to 549 for each of the four molecules in the tetrameric assembly. The previously predicted second and third transmembrane segments (residues 398–414 and 421–441, respectively; Swiss-Prot accession number P08395) are actually buried segments within the major globular domain of $SppA_{EC}$. Thus, we believe these earlier predictions likely hampered creation of soluble constructs of the bacterial signal peptide peptidases by inferring deletion points that spanned crucial folding elements of the enzyme.

The protein fold of $SppA_{EC}$

$SppA_{EC}$ has an α/β protein fold. The major globular domain is made up of three consecutive β–α–β units in a superhelix motif. An extension domain made up of a long α-helix, a long β-strand and two small α-helices sticks out of the major globular domain and creates a dome-like structure in the supermolecular tetrameric complex (Figs. 1 and 2).

The N-terminal (residues 56–316) and C-terminal (residues 326–549) halves of $SppA_{EC}$ are tandem repeats of the above-described protein fold with the residues 317–325 forming a linker between the halves (Fig. 1). Despite only 18% sequence identity between the halves they superimpose with an rmsd of 2.5 Å (Fig. 1b). Each half has nine α-helices and 11 β-strands. The protein/protein interface between the N-terminal and C-terminal halves buries ~1800 $Å^2$ of molecular surface on each half and is linked by 21 hydrogen bonds and one salt bridge. The interface is ~60% nonpolar and the residues that contribute to the interface include the catalytic dyad residues Ser409 and Lys209, the nucleophile Ser409 coming from the C-terminal half and the general base Lys209 arriving from the N-terminal half (Fig. 1). The Ser409 O^{γ} and the Lys209 N^{ζ} are within hydrogen-bonding distance to one another (2.8 Å, average for the four molecules in the asymmetric unit). Besides the Lys209 N^{ζ}, there are no other titratable functional groups within the vicinity of the Ser409 O^{γ} that could play the role of a general base.

The shape, dimensions and surface features of the $SppA_{EC}$ tetramer

Analysis by analytical size-exclusion chromatography in line with multiangle light scattering is consistent with $SppA_{EC}\Delta 2–46$ being a tetramer in solution (data not shown). Previous chemical cross-linking data on the full-length membrane-bound form of $SppA_{EC}$ was consistent with $SppA_{EC}$ being a tetramer.[10] Size exclusion chromatographic analysis of the homologous enzyme from *Arabidopsis thaliana* ($SppA_{AT}$) was also consistent with a tetrameric assembly.[7]

The long extended interface between the SppA monomers (~66 Å × ~27 Å) buries approximately 1450 $Å^2$ of surface (~60% nonpolar) on each monomer. The interface is predominately α-helical in nature but with one long antiparallel two-stranded β-sheet that is formed from the long β-strand in the extension domain that protrudes from the major globular domain in each neighboring monomer. There are a total of 13 hydrogen bonds and one salt bridge between the monomers (Fig. 2).

The $SppA_{EC}$ tetramer is bowl shaped with an opening at its base of approximately 96 Å in diameter (predicted membrane association surface) (Fig. 2). A ridge on the inside of the bowl restricts the opening to approximately 40 Å in diameter. A concave surface is formed between the arched roof of the inverted bowl structure and the ridge (Fig. 2g). This concave surface creates the substrate-binding pockets for four active sites within the $SppA_{EC}$ tetramer. The top of the inverted bowl-shaped structure contains an axial opening approximately 22 Å in diameter that is rimmed with positive charges (Fig. 2a and g). The depth of the bowl-shaped protein is approximately 50 Å from base to axial opening. The wide opening of the bowl has an outer rim of positive charge followed by a broader inner rim of negative charge (Fig. 2e). The molecular surface of the deep interior of the bowl is predominantly nonpolar (Fig. 2e).

The assembly of Ser/Lys dyad active sites from adjacent domains and monomers

It was previously observed that SppA from eubacteria, archaea and thylakoids appears to form two main families of SppA based on size.[5,7,11] The smaller SppA enzymes, which include the characterized Gram-positive *Bacillus subtilis* $SppA_{BS}$[5] and the archaeal *Thermococcus kodakaraensis* $SppA_{TK}$,[6] are approximately half the size of the larger SppA enzymes and are most homologous to the C-terminal half of $SppA_{EC}$. Our structure of $SppA_{EC}$ and our sequence analysis (Fig. 3) bring new insights into the relationship between the smaller and larger forms of SppA. As mentioned above, the N-terminal and C-terminal halves of $SppA_{EC}$ appear to be tandem repeats and the nucleophile Ser409 is located on the C-terminal half, whereas the general

Fig. 1. The protein fold of $SppA_{EC}$. (a) A ribbon diagram of the $SppA_{EC}$ monomer. The N-terminal half of the molecule is shown in blue (residues 56–316), and the C-terminal half of the molecule is shown in red (residues 326–549). The short segment of residues that links the N-terminal and C-terminal regions is shown in yellow (317–325). The positions of the lysine general base (K209) and the serine nucleophile (S409) are labeled and shown as spheres. (b) A superposition of the N-terminal region of $SppA_{EC}$ (blue, residues 56–316) on the C-terminal region (red, residues 326–549). The position of the general base lysine (209) superimposes on A461, which is at the same position as the proposed general base in the archaeal *T. kodakaraensis* $SppA_{TK}$ (K214).[6]

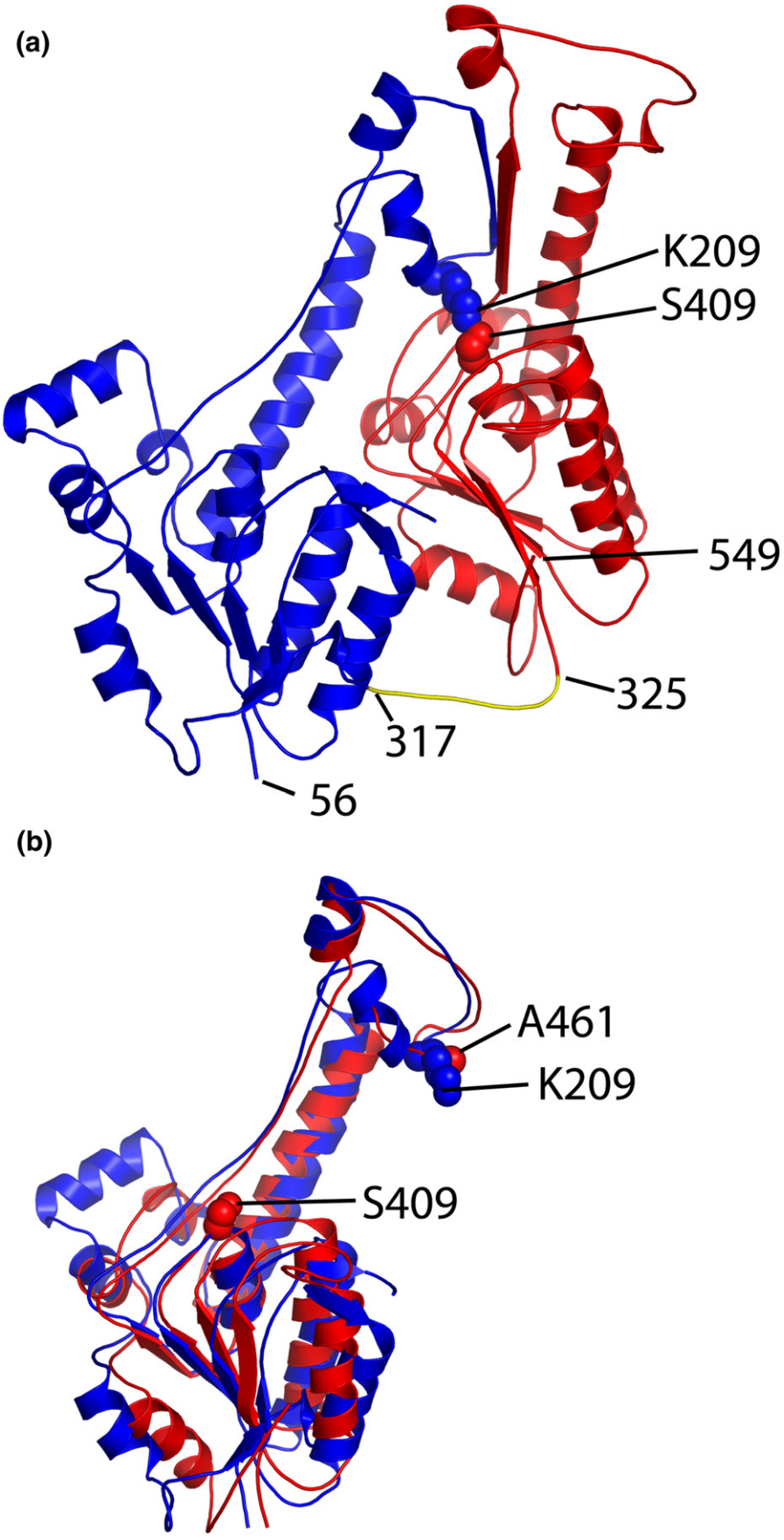

Fig. 1 (*legend on previous page*)

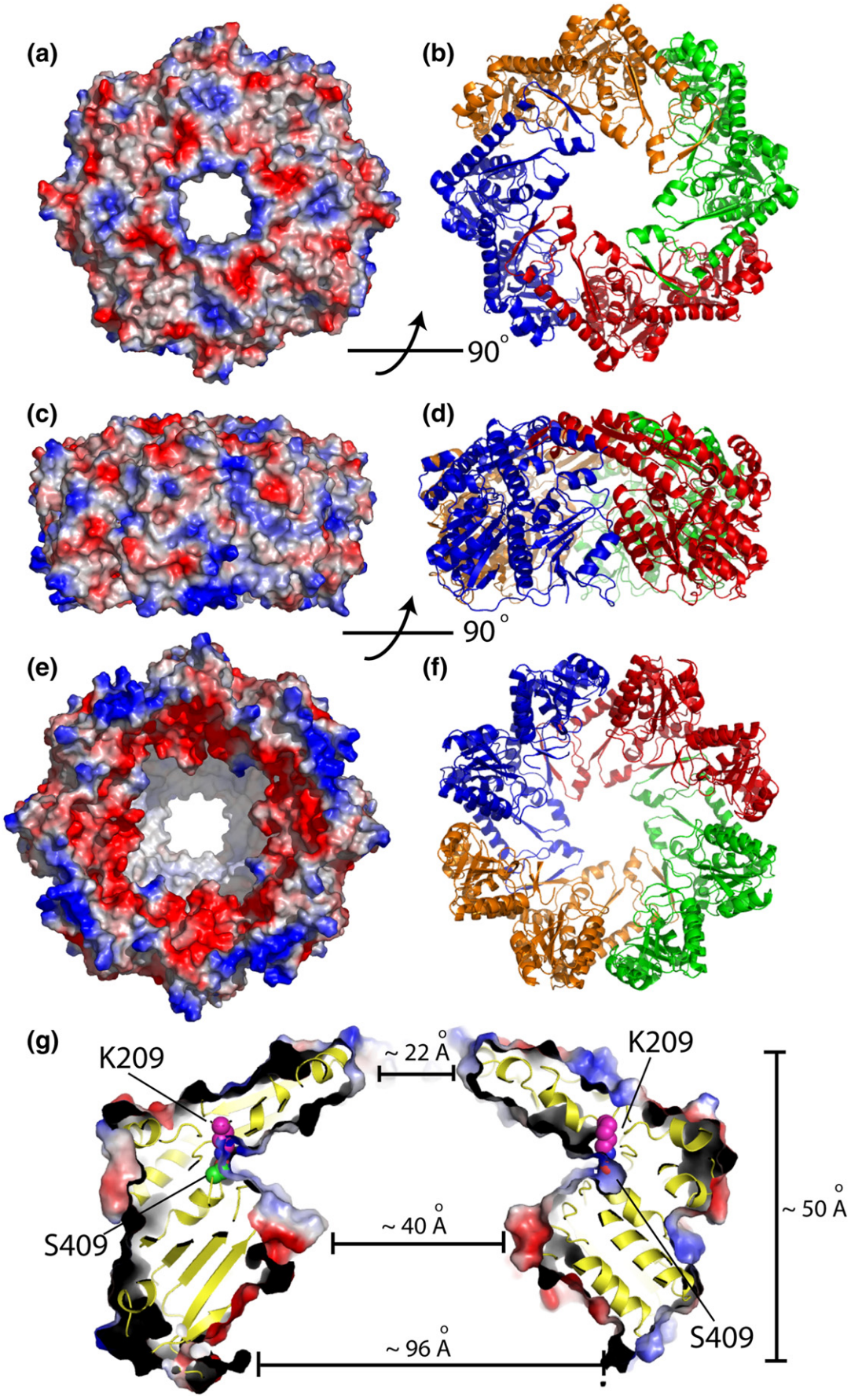

Fig. 2 (*legend on next page*)

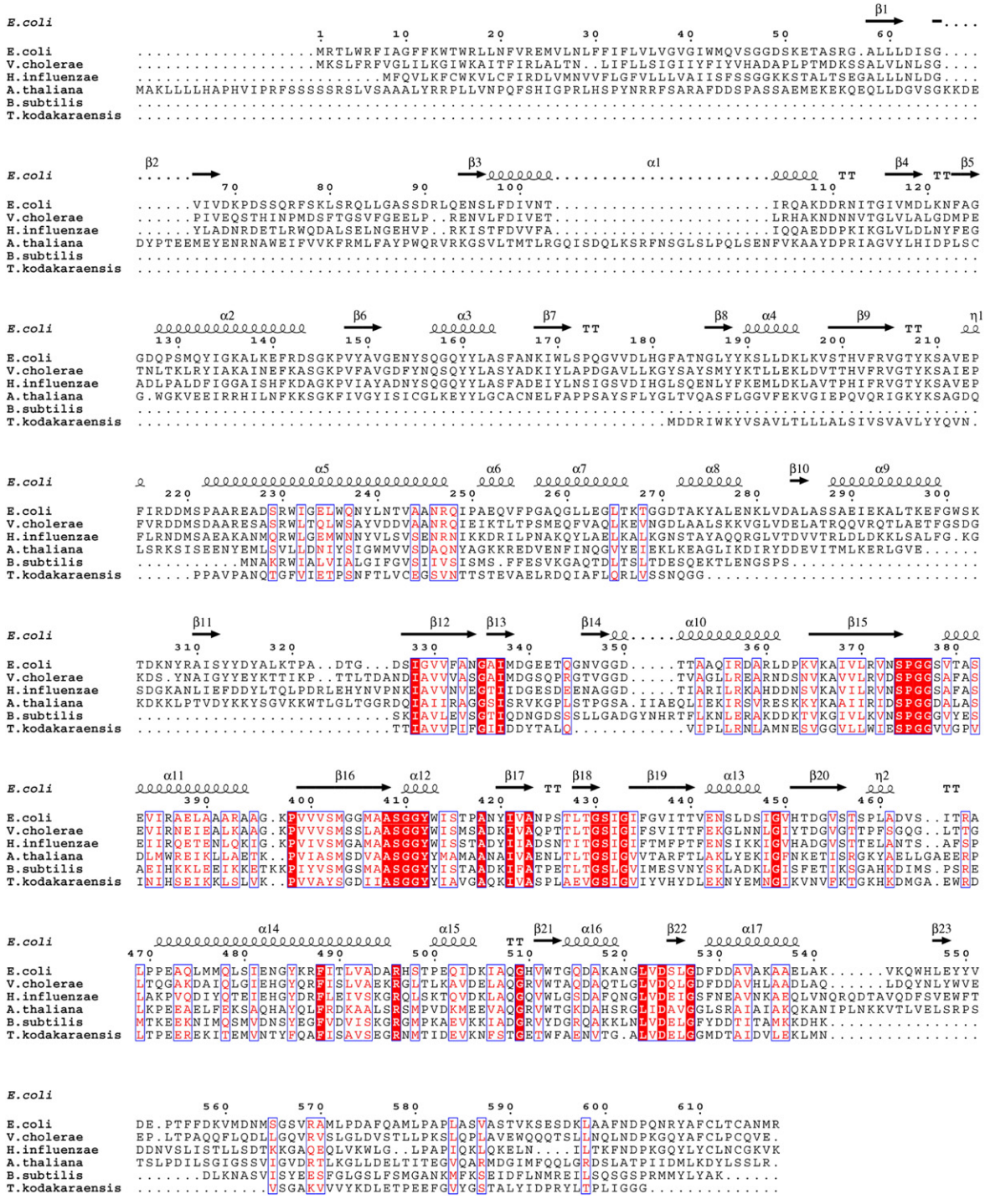

Fig. 3. A sequence alignment of signal peptide peptidases SppA. The secondary structure as calculated by DSSP[12] is shown above the alignment. The sequences were acquired from the Swiss-Prot database with the accession numbers for each sequence in parenthesis: *E. coli* $SppA_{EC}$ (P08395); *Vibrio cholerae* $SppA_{VC}$ (Q9KQK4); *Haemophilus influenzae* $SppA_{HI}$ (P45243); *A. thaliana* $SppA_{AT}$ (Q9SEL8); *B. subtilis* $SppA_{BS}$ (O34525); *T. kodakaraensis* $SppA_{TK}$ (Q5JE91).

base Lys209 is located on the N-terminal half. As can be seen from Fig. 1a, the side-chain functional groups of these residues come together, within hydrogen-bonding distance, at the interface between the N-terminal and C-terminal halves. When the halves are superimposed, the Ser409

Fig. 2. The oligomeric state, electrostatic surface properties and dimensions of $SppA_{EC}$. (a), (c) and (e) show the electrostatic properties of $SppA_{EC}$ mapped on its molecular surface (white, neutral; blue, positive; red, negative). (b), (d) and (f) show different views of ribbon diagrams of $SppA_{EC}$. Each monomer of the tetramer is a different color. (g) A section through the middle of $SppA_{EC}$ revealing the dimensions of the opening to the bowl-shaped structure. The catalytic residues S409 (green) and K209 (magenta) are shown as van der Waals spheres.

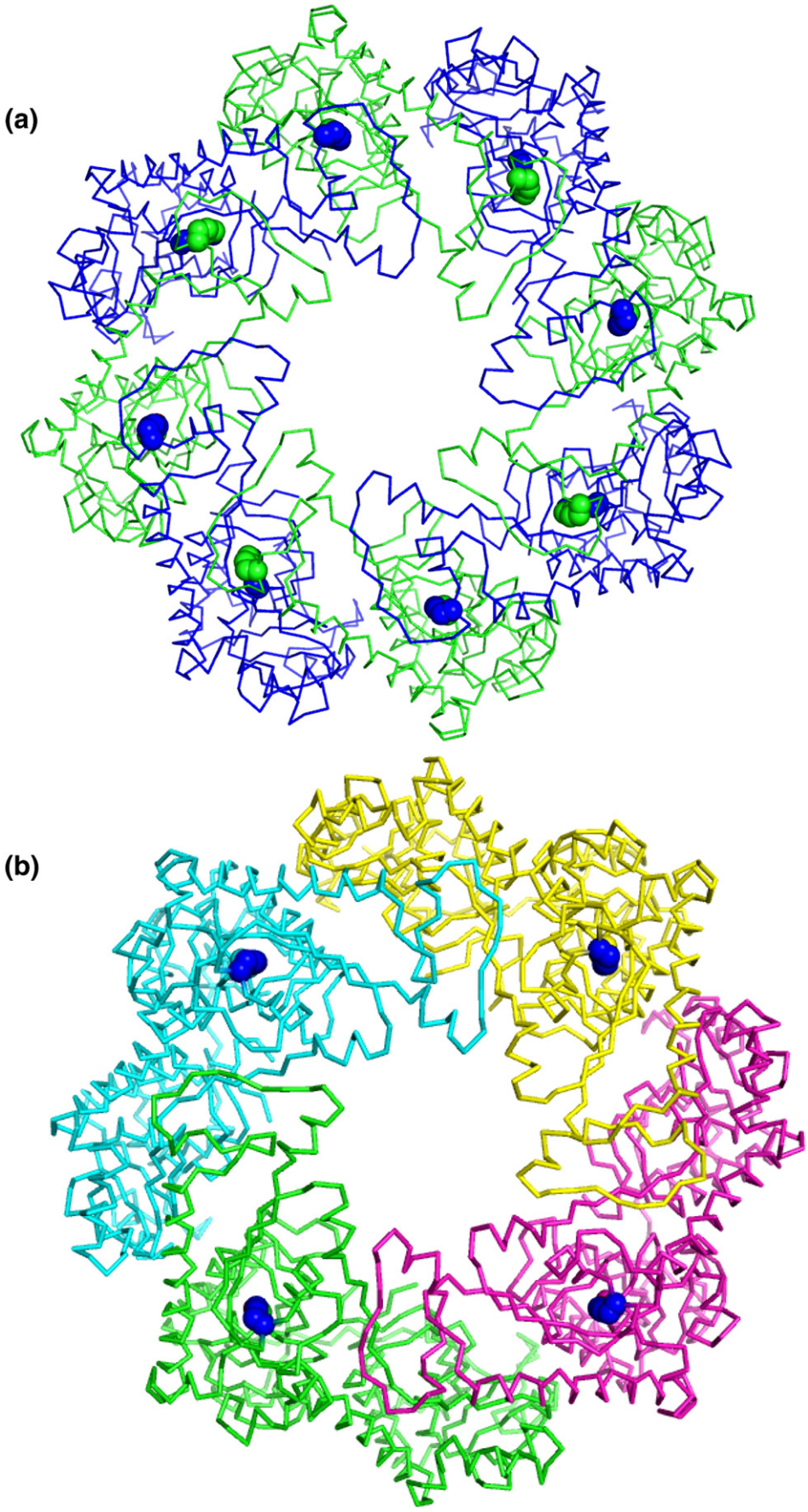

Fig. 4 (*legend on next page*)

nucleophile and Lys209 general base are actually quite far apart (Fig. 1b). Interestingly, this superposition shows that Lys209 from the N-terminal half overlaps well with Ala461 from the C-terminal half, the region most similar to the "half-sized" archaeal SppA. The sequence alignment in Fig. 3 reveals that Ala461 aligns with Lys214 of *T. kodakaraensis* $SppA_{TK}$. Previously, Matsumi *et al.* proposed, based on their mutagenesis work on $SppA_{Tk}$, that Lys214 was the general base in $SppA_{Tk}$.[6] Therefore, together these results are consistent with the position of the nucleophile (Ser409) and general base (Lys209) as seen in the $SppA_{EC}$ structure and provide a structural context for the previous $SppA_{Tk}$ mutagenesis results. In the case of $SppA_{Tk}$, both nucleophile (Ser162) and general base (Lys214) reside on the same domain (or "half"), which would have the effect of putting these residues quite distant from each other, unless they assembled into an oligomeric assembly similar to that seen in $SppA_{EC}$. Only when the oligomeric assembly comes together would the serine nucleophile come into close proximity to the lysine general base. This would suggest that the smaller SppA enzymes form an octameric assembly that would be similar in architecture to the tetrameric assembly seen in the larger ("double sized") SppA relatives, such as $SppA_{EC}$. This also has the effect of doubling the number of active sites in the assembly of the smaller SppA enzymes (Fig. 4). Previous size-exclusion chromatography analysis of $SppA_{TK}$ is consistent with it being an octamer.[11]

Comparison of a Ser/His/Asp protease and a Ser/Lys protease that utilize a similar protein scaffold

Despite limited sequence identity, the N- and C-terminal regions of the Ser/Lys protease $SppA_{EC}$ monomers are similar in protein fold to the monomers seen in the cytoplasmic "classical" Ser/His/Asp ClpP protease. The $SppA_{EC}$ N-terminal or C-terminal halves can be individually superimposed on a ClpP monomer with an rmsd of 3.0 Å (15.6% identity for 147 aligned residues) and 1.8 Å (25% identity for 155 aligned residues), respectively (Fig. 5a and b). ClpP is an ATP-dependent, general protease that works with the chaperones ClpA and ClpX to recognize, unfold and cleave misfolded proteins and transient proteins in the bacterial cytoplasm.[14] Like SppA, ClpP protease also forms a large circular assembly with an axial port hole but the packing of the monomers differs from that of SppA such that a completely unique oligomeric assembly in terms of orientation, size, symmetry and exposed molecular surfaces is created (Figs. 6 and 7). In the case of ClpP, the axial hole is much smaller (~13.5 Å, Fig. 7a and b) and is formed by the globular portion of the structure (Fig. 6a), with the internal sequence extension (or "handle") creating the intercalating components that allow for two seven-fold, bowl-shaped structures to form a 14-mer spherical assembly (Fig. 6c). Additionally, each monomer of ClpP protease encodes a complete active site involving a Ser/His/Asp catalytic triad, again differing from SppA, which utilizes residues at the interface of its tandemly repeated domains to create the Ser/Lys active-site dyad (Fig. 5). The internal surface of the ClpP protease ball-shaped complex that contains the 14 active sites is much more polar than that seen near the 4 active sites in the $SppA_{EC}$ bowl-shaped complex (Fig. 7).

Remarkably, when a ClpP monomer is superimposed onto the C-terminal half of a complete $SppA_{EC}$ monomer (Fig. 5c), the chemically important atoms within the Ser/His/Asp catalytic triad residues of ClpP superimpose with the catalytic atoms within the Ser/Lys catalytic dyad of $SppA_{EC}$ (Fig. 5d). Specifically, the O^{γ} of the serine nucleophiles superimpose as do the $N^{\varepsilon 2}$ of the histidine general base in ClpP and the N^{ζ} of the lysine general base of $SppA_{EC}$, even though the positions of these general base residues come from sequentially and structurally disparate regions of the molecules (Fig. 5d). As is seen in other Ser/Lys proteases,[15] in replacement of the aspartic acid observed in ClpP there is a second serine (Ser431) in SppA that helps coordinate the lysine general base (Fig. 5d). This residue is also conserved across the SppA enzymes (Fig. 3).

The structure of ClpP has been solved to 1.9 Å resolution in complex with a peptide-based inhibitor.[13] Because of the similarity in the substrate-binding architecture in ClpP and $SppA_{EC}$, it is possible to model a peptide substrate into the binding site of $SppA_{EC}$ based on the structure of the ClpP peptide–inhibitor complex (Fig. 8).

The model clearly shows the P1 side chain (modeled as an alanine for clarity) points into the predominantly hydrophobic S1 substrate specificity pocket. The P2 side chain is pointing out towards solvent, away from the binding site, and the P3 side chain is pointing towards the large S3 substrate specificity pocket, which has a negatively charged

Fig. 4. A comparison between the smaller SppA and larger SppA enzymes. (a) Model of the oligomeric assembly proposed for the family of smaller SppA enzymes (such as $SppA_{TK}$) that contain both a serine nucleophile and lysine general base within a single domain (half-sized region as compared to the larger SppA enzymes such as $SppA_{EC}$). The monomers are shown in a C^{α} trace (green and blue in alternating monomers) with the serine nucleophile and lysine general base shown as spheres. This model is consistent with the smaller SppA enzymes, such as *B. subtilis* $SppA_{BS}$ and *T. kodakaraensis* $SppA_{TK}$, having an octameric assembly, each monomer with a serine nucleophile and a lysine general base and forming eight separate active sites. (b) The larger SppA enzymes such as *Escherichia coli* $SppA_{EC}$ have a duplication of the protein fold with one domain containing the serine nucleophile and the other domain containing the lysine general base, forming four separate active sites in the tetrameric assembly.

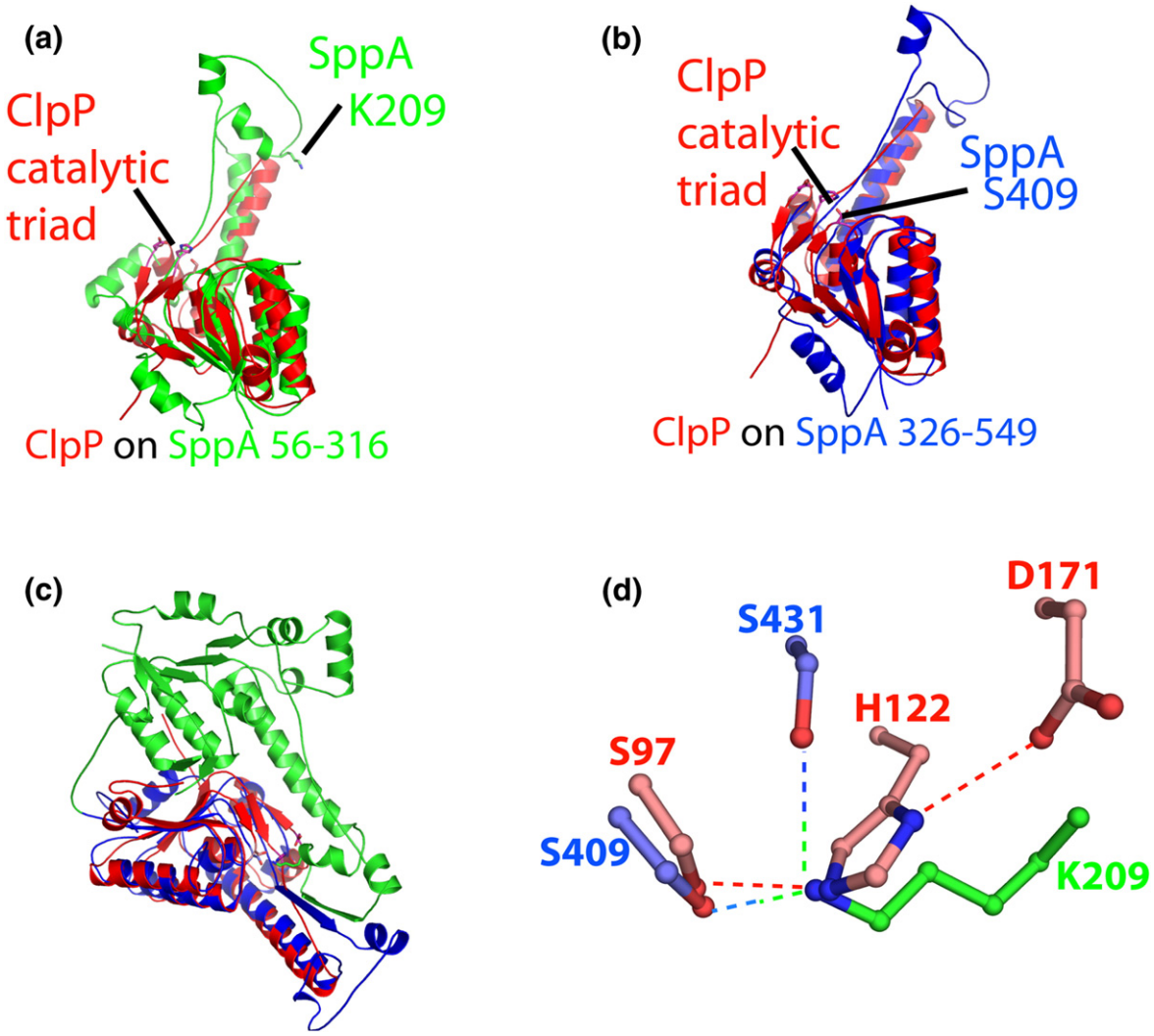

Fig. 5. Comparison of the Ser/Lys protease SppA to the Ser/His/Asp protease ClpP. Superposition of a ClpP monomer (red) on the N-terminal region of SppA (green, residues 56–316) (a) and on the C-terminal region of SppA (blue, residues 326–549) (b). The position of the catalytic triad in ClpP is labeled. The position for the serine nucleophile in the C-terminal domain of SppA overlaps with the position of the serine nucleophile in ClpP (b), but the position of the general base in the N-terminal domain in SppA (K209) is distant from the position of the base (H122) seen in ClpP protease (a). (c) The superposition of ClpP (red) on the full SppA protein (green, blue). (d) A ball-and-stick rendering of the Ser/Lys catalytic dyad active site of $sppA_{EC}$ superposed on the Ser/His/Asp catalytic triad active site of ClpP protease.[13]

surface at the bottom of the pocket. The two aliphatic side chains of residues Val379 and Ile434 form a ridge that corresponds to the dividing point between the S1 and S3 binding pockets. This is similar to what is seen in *E. coli* type I signal peptidase where the analogous residues have been shown to contribute to substrate specificity.[16] The hydrophobic S1 pocket is consistent with previous analysis that showed a preference for substrates with aliphatic residues at the P1 position.[2,4,17] It is possible that the relatively large and negatively charged S3 pocket may interact with the positively charged N-region of the signal peptide. Given these structural insights, it would be interesting to reinvestigate the substrate specificity of SppA.

The model is also consistent with the NH of Gly377 and Gly410 forming the $SppA_{EC}$ oxyanion hole by contributing hydrogen bond donors to the developing oxyanion during the transition state of the hydrolytic reaction. These two glycine residues are completely conserved (Fig. 3). The oxyanion holes in SppA and ClpP are unique among serine proteases in that they use a main-chain amide NH from the residue following the nucleophile rather than the main-chain NH of the nucleophile itself along with the penultimate residue to the nucleophile.[18]

SppA orientation relative to the membrane surface and substrate presentation

$SppA_{EC}$ contains a proposed N-terminal transmembrane segment (residues 29–45) that likely tethers it to the cytoplasmic membrane (Fig. 9a). The N-terminus of our $SppA_{EC}$ structure (residue

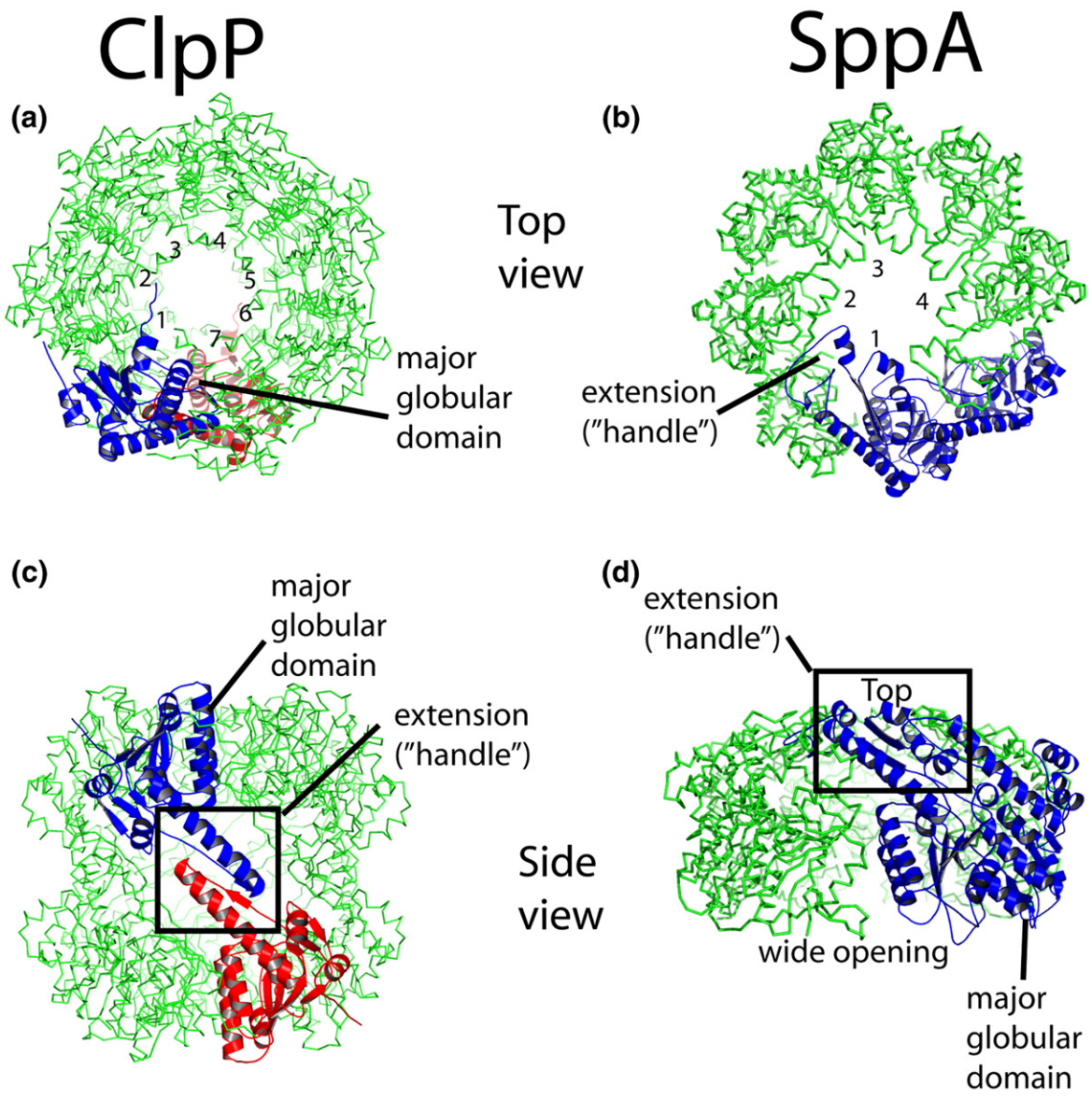

Fig. 6. A comparison of the ClpP protease and the SppA protease oligomeric assembly in terms of monomer orientation, protein complex size and symmetry. (a) and (b) show the comparison between the sevenfold ClpP and the fourfold SppA from the top end view. In ClpP, the major globular domain of the protein forms the axial hole, whereas in SppA, the extension or handle domain forms the axial hole. (c) and (d) show a comparison from the side of the protein complexes. One can see that in ClpP the handle extension domain makes the intercalating interactions between the 7-mer halves of the ClpP ball-shaped 14-mer complex, where as in SppA the handle extension forms the axial hole.

56) is part of the globular domains that forms the rim of the bowl-shaped structure. Therefore, we propose $SppA_{EC}$ is oriented relative to the membrane surface such that its axial hole is pointing away from the inner membrane and the large opening of the bowl is associating with the membrane surface (Fig. 9a). Bioinformatic analysis and the positive inside rule[19] are consistent with the C-terminal catalytic region facing the periplasmic or *trans* side of the inner membrane. Interestingly, surface analysis of $SppA_{EC}$ reveals that the largest depression (yellow surface in Fig. 9b) on the $SppA_{EC}$ molecular surface corresponds to the association interface between individual monomers on the inside of the bowl. Figure 9b shows the interface between two SppA monomers, one monomer in blue and the other in tan and rendered as a ribbon. These depressions lead up to the substrate-binding sites (labeled S1 and S3 in Fig. 9b). It is possible that these depressions on the surface help guide the signal peptide substrate towards the active site (red surface in Fig. 9b, the blue surface corresponds to the P1′ side of the binding site) with the ring of negative charge near the base of SppA attracting the positively charged N-region of the signal peptide. Additional experiments will be needed to address the question of how signal peptides are pulled from the membrane and presented to the SppA-binding site. It is possible that

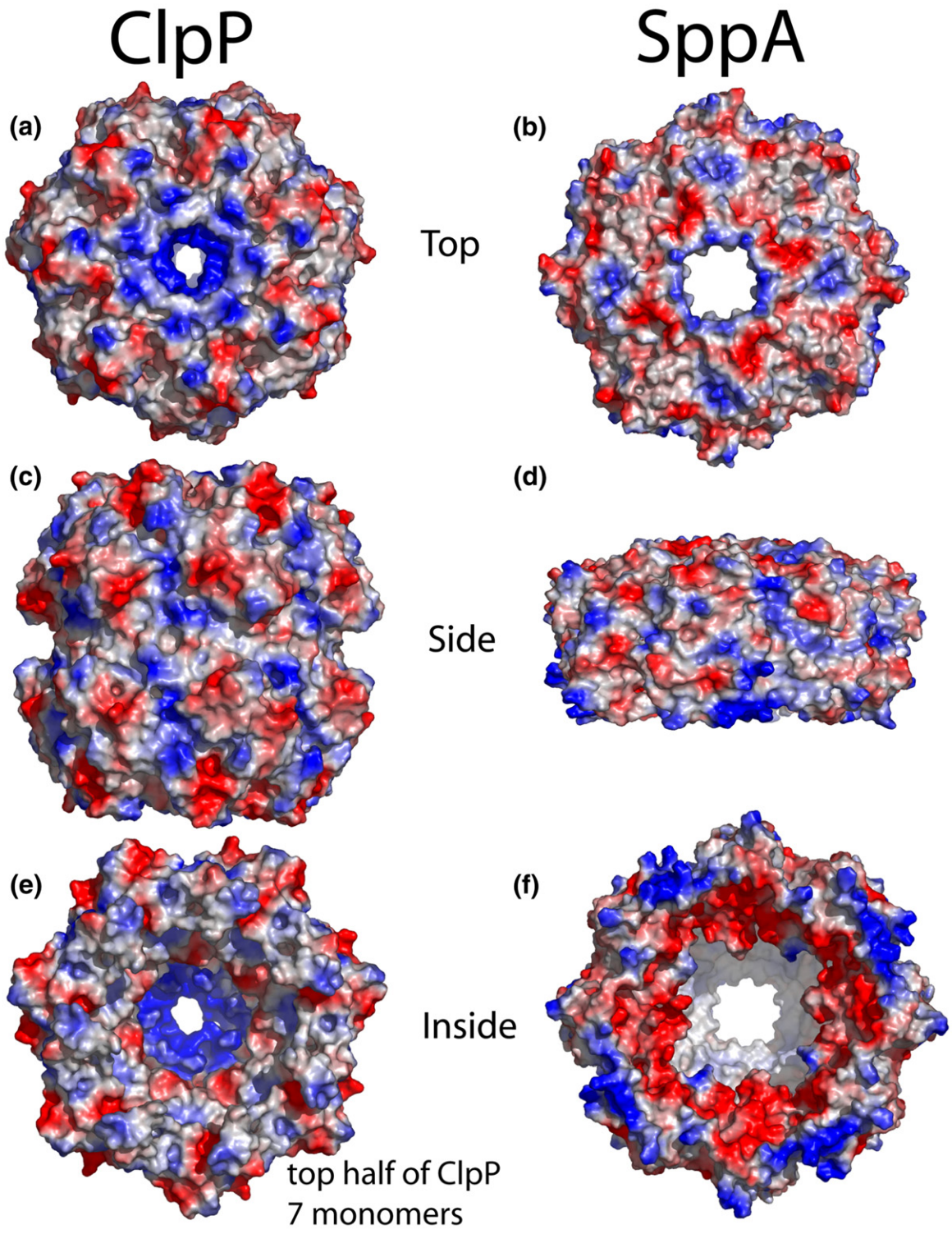

Fig. 7. A comparison of the ClpP protease and the SppA protease oligomeric assembly in terms of the electrostatic nature of the exposed molecular surfaces. (a) and (b) show the difference in the size and charge at the axial hole. (c) and (d) show the difference in the dimensions between the ball-shaped ClpP protease 14-mer (two 7-mer oligomers combined) and the tetrameric SppA that forms a bowl-shaped structure. (e) and (f) show that SppA is much more hydrophobic in the interior of the proteolytic chamber.

SppA may work in concert with other protein(s) that may help facilitate the extraction of the signal peptide from the membrane for subsequent SppA-mediated hydrolysis. This would be reminiscent of the ClpAP and ClpXP chaperone/protease complexes, where ClpX and ClpA are the ATP-dependent chaperones that are required to unfold proteins in the cytosol for hydrolysis by the ClpP pro-

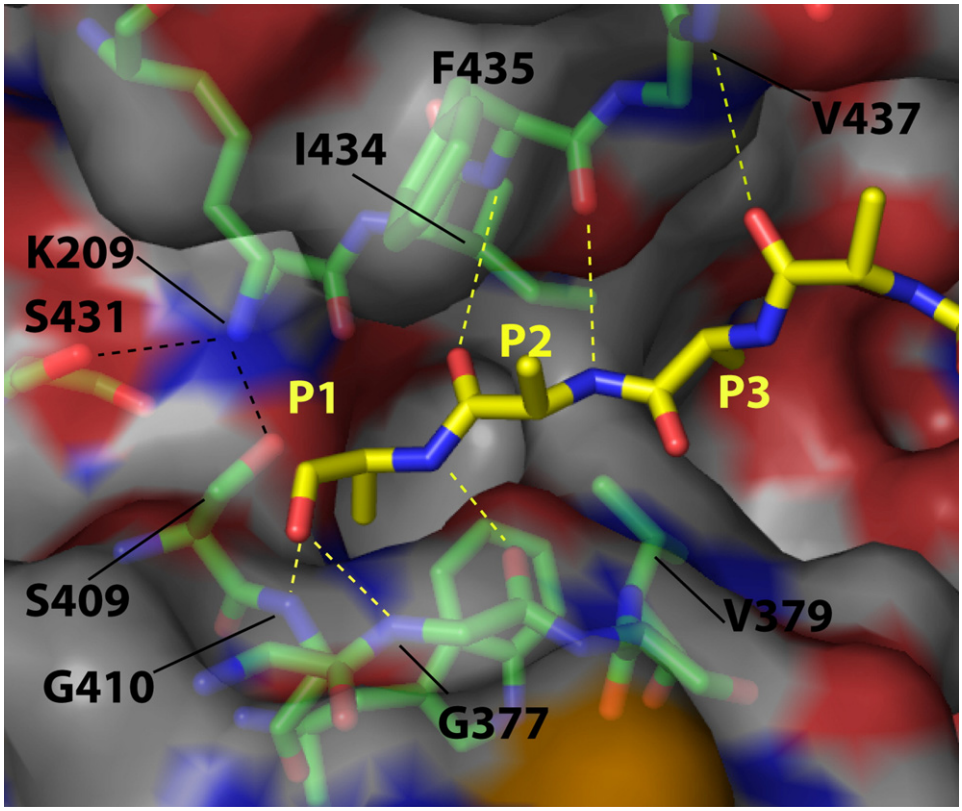

Fig. 8. A model of a peptide (yellow, carbon; blue, nitrogen; red, oxygen) docked into the binding site of $SppA_{EC}$ based on the structure of ClpP with a bound peptide-based inhibitor.[13] The $SppA_{EC}$ binding site residues (green, carbon; blue, nitrogen; red, oxygen) are shown as sticks behind a semitransparent surface colored according to atoms (white, carbon; blue, nitrogen; red, oxygen). Potential hydrogen bonds are shown as dashed lines.

tease.[20,21] The protein architectural similarity in the protein fold of SppA protease and ClpP protease begs the question: Does SppA have a possible role in protein quality control at the membrane surface similar to what ClpP protease performs in the cytoplasm?

Experimental Procedures

Cloning and mutagenesis

The gene *sppA* was cloned out of *E. coli* genomic DNA (K12) using the oligonucleotides: forward primer, 5′-CCA TAT GGG TGG TGA TTC GAA AGA AA-3′; reverse primer, CTC GAG TTA ACG CAT GTT GGC GCA G-3′, which contain the restriction sites *Nde*I and *Xho*I for cloning. The PCR product was cloned into the expression vector pET-28a (Novagen). The final SppA construct (SppAΔ2–46) has the first 20 residues containing the 6xHis affinity tag and the thrombin cleavage site: MGSS-HHHHHH-SSGLVPR-GSH-sppA followed by a Met and then the *sppA* gene starting at the codon for residue Gly47. DNA sequencing confirmed the *sppA* sequence reported in the Swiss-Prot database (accession number P08395). The plasmid pET28-sppAΔ2–46 was transformed into BL21 (DE3) *E. coli* cells. The expressed protein ($SppA_{EC}$Δ2–46) is 593 residues in length (including the 6xHis affinity tag and the thrombin cleavage site) and has a molecular mass of 63,849 Da and a theoretical isoelectric point of 5.9.

Expression and purification of sppA proteins

Overnight small-scale cultures were diluted (2:100 ratio) into LB media containing kanamycin (25 μg/mL), grown at 37 °C for 3.5 h and then induced with IPTG (0.75 mM final concentration) at 27.5 °C for 3.5 h. Cells were pelleted by centrifugation and then lysed using the Avestin Emulsiflex-3C cell homogenizer. The lysate was clarified by centrifugation (37,000*g*) for 20 min. The supernatant was applied to a nickel NTA column (5 mL column volume, Qiagen) equilibrated with 5 mM imidazole, 50 mM Tris–HCl, pH 7.5, and 100 mM NaCl. The lysate was applied three times to the column and then the column was washed with 50 mM imidazole, 50 mM Tris–HCl, pH 7.5, 100 mM NaCl (10 column volumes) and then finally eluted with a step gradient (100 to 500 mM imidazole with 100 mM increments). The protein eluted from the column between 200 and 400 mM imidazole. The protein was then concentrated *via* an Amicon ultracentrifuge filter (Millipore) and applied to a Sephacryl S-100 HiPrep 26/60 size-exclusion chromatography column on an AKTA Prime system (Pharmacia Biotech) run at 1 mL/min, with a buffer containing 50 mM Tris–HCl, pH 7.5, 100 mM NaCl, and 1 mM ethylenediaminetetraacetic acid. Fractions containing pure sppA were analyzed by SDS-PAGE and concentrated to 8 mg/mL. Analytical size-exclusion chromatography in line with multiangle light-scattering analysis is consistent with $SppA_{EC}$Δ2–46 being a monodispersed tetramer in solution in the absence of detergent (data not shown). Cleavage assays show that $SppA_{EC}$Δ2–46 is catalytically active (data not shown). The

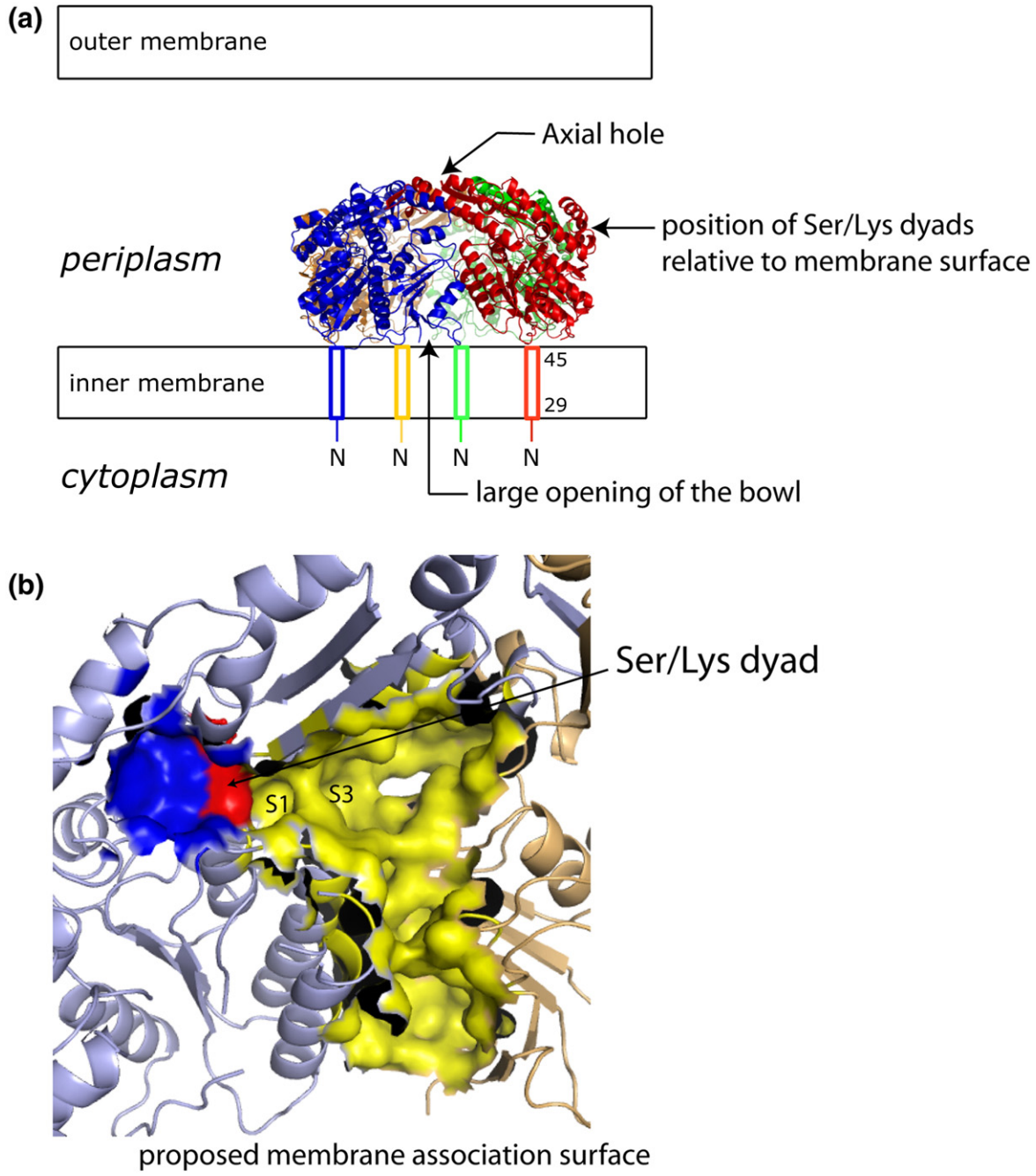

Fig. 9. The orientation of $SppA_{EC}$ with respect to the membrane surface. (a) The tetrameric $SppA_{EC}$ is shown in a ribbon diagram with each monomer a different color. The proposed position for the N-terminal transmembrane segment (residues 39–45) for each monomer (which is missing from the construct that was used for crystallization) is schematically shown. The position of the axial hole, the large opening of the bowl that leads to the active sites and the position of the catalytic residues relative to the membrane surface are labeled. The diagram is not to scale. (b) The interface between each SppA monomer creates the largest pocket on the SppA surface and a continuous path between the proposed membrane association surface and the active site. The interface between two SppA monomers is shown with each monomer (blue and tan) rendered in ribbon. The molecular surface of the pocket formed between the monomers is shown (yellow surface) and the position of the active site (Ser/Lys dyad in red surface) and the proposed membrane association surface is labeled. The S1 and S3 substrate-binding pockets are labeled and the S1′ side of the binding site is shown as blue surface.

selenomethionine (Se-Met)-incorporated $SppA_{EC}\Delta 2–46$ was prepared by making an overnight culture of BL21 (DE3) cells containing pET28-sppAΔ2–46 in M9 minimal media (10 mL of overnight culture was used to inoculate 1 L of M9 minimal media). All growth media were supplemented with 25 μg/mL kanamycin for selection. Cultures were grown for 8 h at 37 °C with shaking and a mixture of the following L-amino acids were added directly: 100 mg lysine, phenylalanine, threonine; 50 mg isoleucine, leucine, valine; 60 mg Se-Met. After 15 min, protein expression was induced by the addition of 0.75 mM IPTG (final concentration). The purification procedure was the same as that used for the native protein.

Crystallization

The crystals used for SAD data collection were grown by the hanging-drop vapor-diffusion method. The crystallization drops were prepared by mixing 1 μL of protein (8 mg/mL) with 1 μL of reservoir solution and then equilibrating the drop against 500 μL of reservoir solution. The $SppA_{EC}\Delta 2–46$ protein produced crystals in the space group $P2_1$ with unit cell dimensions of 93.0 Å×153.0 Å×100.2 Å, β=104.2° (Se-Met), and 94.2 Å×153.5 Å×100.7 Å, β=104.3° (native), with four molecules in the asymmetric unit and a Matthews coefficient of 2.7 $Å^3$ Da^{-1} (54.6% solvent). The optimized crystallization reservoir condition was 100 mM Tris–HCl (pH 7.5), 18% polyethylene glycol 3350, 200 mM K_2HPO_4. The crystallization was conducted at 18 °C. The cryosolvent condition was the same as that of the reservoir solution but with 20% of the water replaced with glycerol. Crystals were incubated in the cryosolvent condition for approximately 5 min before being flash-cryocooled in liquid nitrogen.

Data collection

Diffraction data were collected on Se-Met-incorporated crystals at beamline X4A of the Brookhaven National Laboratory, National Synchrotron Light Source, using a Q4 CCD detector and a Huber Eulerian Goniometer. The crystal-to-detector distance was 175 mm. Data were collected with 1° oscillations, and each image was exposed for 15 s. The native crystals were collected at the Advanced Light Source, Lawrence Berkeley National Laboratory, University of California at Berkeley, beamline 8.2.1 using an ADSC Q315R detector. The crystal-to-detector distance was 340 mm. Data were collected with 1° oscillations, and each image was exposed for 5 s. The diffraction data were processed with the program HKL2000.[22] See Table 1 for data collection statistics.

Structure determination and refinement

The $SppA_{EC}$ structure was solved by SAD using a data set collected at the peak wavelength (0.97925 Å), the programs HKL2MAP[23] version 0.2 and AUTOSOL within PHENIX[24] version 1.3.[24] The program AUTOBUILD within PHENIX version 1.3[24] automatically constructed approximately 80% of the polypeptide chain and performed density modification.[24] The rest of the model was built and fit using the program COOT.[25] The structure was refined using the program REFMAC5.[26] The final models were obtained by running restrained refinement in REFMAC5 with TLS restraints obtained from the TLS motion-determination server.[27] The native structure was solved by molecular replacement using chain A from the SAD structure as a search model and then refined as described above. The data collection, phasing and refinement statistics are summarized in Table 1.

Structural analysis

The secondary structural analysis was performed with the programs DSSP,[12] HERA[28] and PROMOTIF.[29] The programs SUPERIMPOSE[30] and SUPERPOSE[31] were used to overlap coordinates for structural comparison. The program CONTACT within the program suite CCP4 was used to measure the hydrogen bonding and van der Waals contacts.[32] The program CASTp[33] was used to analyze the molecular surface and measure the substrate-binding site.

Table 1. Data collection, phasing and refinement statistics

	Se-Met	Native
Crystal parameters		
Space group	$P2_1$	$P2_1$
a, *b*, *c* (Å)	93.0, 153.0, 100.2	94.2, 153.5, 100.7
α, β, γ (°)	β=104.2	β=104.3
Data collection statistics		
Wavelength (Å)	0.97925	1.0000
Resolution (Å)	50.0–2.8 (2.9–2.8)	50.0–2.5 (2.6–2.5)
Total reflections	511,503	626,825
Unique reflections	69,169 (5523)	86,579 (7232)
R_{merge}[a]	0.178 (0.450)	0.122 (0.324)
Mean $(I)/\sigma(I)$	15.6 (3.3)	16.0 (3.2)
Completeness (%)	97.8 (78.0)	96.2 (80.8)
Redundancy	7.4 (6.1)	7.2 (6.0)
Phasing statistics		
No. of sites	50 (out of a possible 60)	–
Overall FOM (50–3.1 Å)	0.32	–
Overall FOM (after density modification)	0.74	–
Refinement statistics		
Protein molecules (chains) in A.U.	4	4
Residues	1976	1986
Water molecules	474	554
Total no. of atoms	15,134	16,266
R_{cryst}[c]/R_{free}[d] (%)	20.5/25.8	21.5/25.2
Average *B*-factor ($Å^2$) (all atoms)	27.3	24.1
rmsd on angles (°)	1.839	1.765
rmsd on bonds (Å)	0.017	0.016

Values in brackets are for the highest resolution shell.

The program SURFACE RACER 1.2[34] was used to measure the solvent-accessible surface of the protein and individual atoms within the protein.[34] A probe radius of 1.4 Å was used in the calculations. The Protein–Protein Interaction Server[35,36] was used to analyze the interactions between the molecules in the asymmetric unit. The stereochemistry of the structures was analyzed with the program PROCHECK.[37] The DALI server was used to find proteins with similar protein folds.[38]

Figure preparation

Figures were prepared using PyMol.[39] The alignment figure was prepared using the programs CLUSTALW[40] and ESPript.[41]

Protein Data Bank accession codes

Coordinates have been deposited with accession code 3BEZ for the Se-Met data and 3BFO for the native data.

Acknowledgements

We thank Dr. R. M. Sweet and all the RapiData 2007 staff at the Brookhaven National Laboratory

NSLS beamline X4A. This work was supported in part by a Canadian Institute of Health Research operating grant, a National Science and Engineering Research Council of Canada discovery grant, a Michael Smith Foundation for Health Research Senior Scholar award, a Canadian Foundation of Innovation grant (to M.P.) and a Michael Smith Foundation for Health Research Postdoctoral Fellow Award (to D.C.O.).

References

1. Blobel, G. & Sabatini, D. D. (1971). Ribosome–membrane interaction in eukaryotic cells. *Biomembranes*, **2**, 193–195.
2. Novak, P. & Dev, I. K. (1988). Degradation of a signal peptide by protease IV and oligopeptidase A. *J. Bacteriol.* **170**, 5067–5075.
3. Miller, C. G. (2004). Protease IV. In *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D. & Woessner, J. F., eds), pp. 2039–2041, 2nd edit. Elsevier, London.
4. Pacaud, M. (1982). Purification and characterization of two novel proteolytic enzymes in membranes of *Escherichia coli*. Protease IV and protease V. *J. Biol. Chem.* **257**, 4333–4339.
5. 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.
6. 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.
7. 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.
8. Rawlings, N. D. & Barrett, A. J. (1999). MEROPS: the peptidase database. *Nucleic Acids Res.* **27**, 325–331.
9. 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.
10. Ichihara, S., Suzuki, T., Suzuki, M. & Mizushima, S. (1986). Molecular cloning and sequencing of the sppA gene and characterization of the encoded protease IV, a signal peptide peptidase, of *Escherichia coli*. *J. Biol. Chem.* **261**, 9405–9411.
11. Matsumi, R., Atomi, H. & Imanaka, T. (2005). Biochemical properties of a putative signal peptide peptidase from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bacteriol.* **187**, 7072–7080.
12. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
13. Szyk, A. & Maurizi, M. R. (2006). Crystal structure at 1.9 A of *E. coli* ClpP with a peptide covalently bound at the active site. *J. Struct. Biol.* **156**, 165–174.
14. Wang, J., Hartling, J. A. & Flanagan, J. M. (1997). The structure of ClpP at 2.3 A resolution suggests a model for ATP-dependent proteolysis. *Cell*, **91**, 447–456.
15. 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.
16. Karla, A., Lively, M., Paetzel, M. & Dalbey, R. E. (2005). The identification of residues that control signal peptidase cleavage fidelity and substrate specificity. *J. Biol. Chem.* **280**, 6731–6741.
17. Pacaud, M. (1982). Identification and localization of two membrane-bound esterases from *Escherichia coli*. *J. Bacteriol.* **149**, 6–14.
18. Menard, R. & Storer, A. C. (1992). Oxyanion hole interactions in serine and cysteine proteases. *Biol. Chem. Hoppe-Seyler*, **373**, 393–400.
19. von Heijne, G. (1989). Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature*, **341**, 456–458.
20. Grimaud, R., Kessel, M., Beuron, F., Steven, A. C. & Maurizi, M. R. (1998). Enzymatic and structural similarities between the *Escherichia coli* ATP-dependent proteases, ClpXP and ClpAP. *J. Biol. Chem.* **273**, 12476–12481.
21. Kessel, M., Maurizi, M. R., Kim, B., Kocsis, E., Trus, B. L., Singh, S. K. & Steven, A. C. (1995). Homology in structural organization between *E. coli* ClpAP protease and the eukaryotic 26 S proteasome. *J. Mol. Biol.* **250**, 587–594.
22. Otwinowski, Z. (1993). Denzo. In *Denzo* (Sawyer, L., Isaacs, N. & Baily, S., eds), pp. 56–62, SERC Daresbury Laboratory, Warrington, UK.
23. Pape, T. & Schneider, T. R. (2004). HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Crystallogr.* **37**, 843–844.
24. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W. *et al.* (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58**, 1948–1954.
25. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
26. Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **57**, 122–133.
27. 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.
28. Hutchinson, E. G. & Thornton, J. M. (1990). HERA—a program to draw schematic diagrams of protein secondary structures. *Proteins*, **8**, 203–212.
29. Hutchinson, E. G. & Thornton, J. M. (1996). PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212–220.
30. Diederichs, K. (1995). Structural superposition of proteins with unknown alignment and detection of topological similarity using a six-dimensional search algorithm. *Proteins*, **23**, 187–195.
31. Maiti, R., Van Domselaar, G. H., Zhang, H. & Wishart, D. S. (2004). SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res.* **32**, W590–W594.
32. Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–763.
33. 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.

34. Tsodikov, O. V., Record, M. T., Jr & Sergeev, Y. V. (2002). Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *J. Comput. Chem.* **23**, 600–609.
35. Jones, S. & Thornton, J. M. (1995). Protein–protein interactions: a review of protein dimer structures. *Prog. Biophys. Mol. Biol*. **63**, 31–65.
36. Jones, S. & Thornton, J. M. (1996). Principles of protein–protein interactions. *Proc. Natl. Acad. Sci. USA*, **93**, 13–20.
37. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
38. Holm, L. & Sander, C. (1998). Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.* **26**, 316–319.
39. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, Palo Alto, CA.
40. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: 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.
41. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.