Supporting Material for:

Structure of signal peptide peptidase A with C-termini bound in the active sites: insights into specificity, self-processing and regulation.

*Sung-Eun Nam and Mark Paetzel**

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

CORRESPONDING AUTHOR: Mark Paetzel, E-mail: mpaetzel@sfu.ca, Telephone: (778)
782-4230 , Fax: (778) 782-5583

Figure S1. Electrospray ionization mass-spectroscopy analysis of $\text{SppA}_{\text{BS}}^{\Delta 1-25}\text{K199A}$ without (A) and with (B) limited proteolysis treatment with thermolysin. **A.** It is well documented that many proteins expressed within the cytoplasm of *Escherichia coli* are processed by methionine aminopeptidase (see references below). The preference for this processing is a small residue side chain (e.g., **Gly**, Ala, Ser, Cys, Pro, Thr, and Val) just after the N-terminal **Met**. The sequence at the N-terminus of the $\text{SppA}_{\text{BS}}^{\Delta 1-25}\text{K199A}$ construct contains an N-terminal six-histidine tag with a linker sequence for thrombin cleavage (MGSSHHHHHHSSGLVPRGSH). The calculated molecular mass of this construct (minus the N-terminal Met) is 35,993.7. $\text{SppA}_{\text{BS}}^{\Delta 1-25}\text{K199A}$ has measured molecular mass of 35,993.80 Da. **B.** Electrospray ionization mass-spectral analysis of thermolysin treated $\text{SppA}_{\text{BS}}^{\Delta 1-25}\text{K199A}$ (sample used for crystallization). N-terminal sequencing of the protein from SppA_{BS} crystals shows that the protein starts at Leu 51. The calculated molecular mass for a SppA_{BS} sequence Leu51-Gly295 is 26,624.1 Da. The measured molecular mass is 26,624.50 Da.

- Xiao,Q., Zhang,F., Nacev,B.A., Liu,J.O. and Pei,D. (2010) Protein N-terminal processing: substrate specificity of *Escherichia coli* and human methionine aminopeptidases. *Biochemistry*. **49**(26), 5588-5599.

- Frottin,F., Martinez,A., Peynot,P., Mitra,S., Holz,R.C., Giglione,C. and Meinnel,T. (2006) The proteomics of N-terminal methionine cleavage. *Mol.Cell.Proteomics*. **5**(12), 2336-2349.

Figure S2. Omit map (3σ) for the last four residues of the SppA_{BS} protomer. The omit ($F_o - F_c$ positive difference electron density) map (green mesh) was created by removing the last four residues, ²⁸⁸ISYE²⁹¹, of SppA_{BS}. On the left, the SppA_{BS} monomer in cartoon (grey) is shown with the omit map for ²⁸⁸ISYE²⁹¹. On the right is a close up view of the omit map with the refined model for ²⁸⁸ISYE²⁹¹ in stick representation (grey = carbon, blue = nitrogen and red = oxygen).

Figure S3. Electron density maps for the SppA_{BS} C-terminal peptide building process. All the $F_o - F_c$ electron density maps (green mesh) are contoured at 3σ and all $2F_o - F_c$ electron density maps (blue mesh) are contoured at the 1σ level. **A.** The circular shape of the positive difference electron density map ($F_o - F_c$) observed in the substrate binding grooves of the octameric SppA_{BS} (shown in grey surface representation). **B.** The positive difference electron density map ($F_o - F_c$) that is observed in each of the eight substrate binding grooves of SppA_{BS}. **C.** A poly-alanine peptide (sticks with oxygen in red, nitrogen in blue, and carbon in green) is fit and refined within the $F_o - F_c$ positive electron density map after the first round of refinement of the eight protomers. (blue mesh = $2F_o - F_c$, green mesh = $F_o - F_c$) **D.** The C-terminus of SppA_{BS} has the sequence ³²⁶SPRMMYLYAK³³⁵. To confirm the presence of tyrosine at position 331 (P1 position), a peptide with this sequence was fit into electron density with Tyr331 mutated to lysine. A zoomed in view, looking in the eye's direction, shows the $2F_o - F_c$ electron density map (blue mesh) for the Y331K mutation as well as the $F_o - F_c$ positive electron density map (green mesh). Note the $F_o - F_c$ positive electron density on the edges of the lysine side chain, consistent with an aromatic ring occupying this density. **E.** A SppA_{BS} C-terminal peptide, ³²⁶SPRMMYLYAK³³⁵, is shown with the surrounding $2F_o - F_c$ map (blue mesh).

Figure S4. LC-MS/MS fragmentation analysis to confirm the identity of the peptide that is bound within the substrate-binding grooves of SppA_{BS}. The top panel is the HPLC chromatogram showing the peptide peak. The middle panel shows the MS/MS fragmentation analysis. The bottom table shows the confirmation of the peptide's identification as that of the C-terminus of SppA_{BS}.

Methods for the LC-MS/MS fragmentation analysis (peptide sequencing) in Figure S4.

Instrumentation: Bruker maXis Impact Quadrupole Time-of-Flight LC/MS System. The system consists of an Agilent 1200 HPLC and a Bruker maXis Impact Ultra-High Resolution tandem TOF (UHR-Qq-TOF) mass spectrometer. Software: Compass 1.5 and BioTools 3.2, Condition for Peptide Sequencing: Mass Spectrometer Ionization Mode: Positive Electrospray Ionization (+ESI), Gas Temp (°C): 200, Gas Flow (l/min): 9, Nebulizer (bar): 4, Capillary Voltage (V): 4500, Collision Energy (eV): 20, Isolation Mass (Da): 672, Isolation Width (Da): 5

HPLC: HPLC Column: Zorbax 300-C8, particle size 3.5 micron, 50 (length) x 2.1 (diameter) mm

HPLC Gradient Table: Solvent A: Water with 0.1% formic acid, Solvent B: Acetonitrile with 0.1 % formic acid

Time (min)	% Solvent B	Flow rate mL/min
0	10	0.4
1	10	0.4
13	99	0.4
14	99	0.4
14.1	10	0.4
20 (stop)	10	0.4

Figure S5. A stereo-view of the SppA_{BS} substrate binding groove residues interacting with its C-terminal peptide. The peptide carbon atoms are shown in green (nitrogen atoms are blue, oxygen atoms are red). The carbon atoms of the binding groove residues that interact with the peptide are shown as grey sticks. Black dashed lines represent the hydrogen bonds. Waters are represented as spheres. The magenta waters bridge peptide atoms and binding groove main chain atoms. The cyan water is conserved within each binding groove but does not interact directly with the peptide.

Figure S6. Electrospray ionization time-of-flight mass-spectroscopy (ESI-TOF-MS) analysis of SppA_{BS}^{Δ2-54}K199A after a 12 day incubation at 23 °C reveals multiple cleavage products. 160 μg of SppA_{BS}^{Δ2-54}K199A was incubated in 100 μl Buffer A at 23 °C. After 12 days, the sample was diluted 10 times in double distilled water and then analyzed by ESI-TOF-MS. The SppA_{BS}^{Δ2-54}K199A in this experiment was not exposed to thermolysin.

Table S1. The measured and calculated molecular masses for the four most abundant peaks observed in the mass spectrometry analysis of self-degraded SppA_{BS}^{Δ2-54}K199A after a 12 day incubation at 23 °C (in Figure S6).

Measured molecular mass (Da)	Calculated molecular mass (Da) ^a	Diff ppm ^b	SppA _{BS} sequences based on Mass-spec ^c
32949.48	32949.4	2.4	<u>GSSHHHHHHSSGLVPRGSH</u> M-Ser55-Lys335
31198.64	31198.6	1.3	<u>GSHM</u> -Ser55-Lys335
27140.92	27140.8	4.4	<u>GSHM</u> -Ser55-Phe300
26332.00	26331.9	3.8	<u>GSHM</u> -Ser55-Glu292

^a, The molecular mass was calculated using the sequence on the right of the table.

^b, Diff ppm is calculated using the following equation: diff ppm = (measured mass – calculated mass)/calculated mass x 10⁶.

^c, The residues that are underlined are those belonging to the N-terminal six histidine affinity tag with the linker (GSSHHHHHHSSGLVPRGSHM).

Figure S7. A standard curve for the size-exclusion chromatography column (Superdex200 column) used to access the oligomeric state of SppA_{BS} C-terminal truncation mutants. The graph is shown with K_{av}, (partition coefficient) on the X-axis and the log of the molecular mass of the standards (Table S2) on the Y-axis. The K_{av} values were calculated using the equation, $K_{av} = (V_e - V_o) / (V_t - V_o)$, where V_t is the column volume (23 mL), V_o is the void volume (8.31 mL) and V_e is the elution volume of each standard or SppA_{BS} construct (Table S2). The measured molecular mass of the sample was calculated using the equation shown in the graph, where x is K_{av} of the protein sample and y is log of the molecular mass of the protein sample.

Table S2. The size-exclusion chromatography elution volumes for the standards and SppA_{BS}^{Δ1-25}K199A constructs and their calculated molecular masses using the standard curve and resulting formula seen in Figure S7.

Standards and SppA constructs	Elution volume (mL)	Molecular mass (kDa) via standard curve
Thyroglobulin (669 KDa)	9.08	
Ferritin (440 KDa)	10.61	
Aldolase (158 KDa)	12.65	
Ovalbumin (43 KDa)	15.04	
SppA _{BS} ^{Δ1-25} K199A Δ295-335 ^a	11.87	209
SppA _{BS} ^{Δ1-25} K199A Δ307-335 ^a	11.87	209
SppA _{BS} ^{Δ1-25} K199A Δ329-335 ^a	11.75	221
SppA _{BS} ^{Δ1-25} K199A ^a	11.80	215

^a Each construct was treated with thermolysin before being applied to the size-exclusion chromatography column. N-terminal sequencing, crystallographic analysis and electrospray ionization mass spectrometry results (Figure S1) are consistent with the thermolysin treatment of SppA_{BS}K199A resulting in a protein that starts and ends at Leu51-Gly295 (calculated mass: 26,624.1 Da). A thermolysin treated SppA_{BS}K199A octamer would then have the molecular mass: 26,624.1 Da x 8 = 212,992.8 Da ~ **213 kDa**).

Figure S8. SDS-PAGE gels showing the fractions collected for the truncated SppA_{BS} constructs after being applied to a size-exclusion chromatography column (Superdex200). **A, B, C** and **D** are fractions collected from thermolysin treated SppA_{BS}^{Δ1-25}K199A $\Delta 295-335$, SppA_{BS}^{Δ1-25}K199A $\Delta 307-335$, SppA_{BS}^{Δ1-25}K199A $\Delta 329-335$, and SppA_{BS}^{Δ1-25}K199A (Full C-terminus). Each fraction is 1 mL in volume. The fraction number as well as the input protein is labeled at the top of each gel. The SDS-PAGE gel molecular mass standards are labeled on the left side of each gel. The elution chromatogram for each construct is shown in Figure 7.