Structural insights into the role of the N-terminus in the activation and function of extracellular serine protease from *Staphylococcus epidermidis*

Kartik Manne and Sthanam V. L. Narayana*

Received 27 August 2019
Accepted 7 November 2019

Edited by Z. S. Derewenda, University of Virginia, USA

Keywords: extracellular serine protease; glutamyl endopeptidase; zymogen; crystal structure; *Staphylococcus epidermidis*.

PDB references: extracellular serine protease, active-site serine mutant, 6pym; zymogen mutant, S66V, 6q12; zymogen mutant, S235A, 6q24; N-terminal locked with one pro-peptide residue, V67C, D255C, 6tya; N-terminal locked with eight pro-peptide residues, V67C, D255C, 6u1b

Supporting information: this article has supporting information at journals.iucr.org/d

Center for Biophysical Sciences and Engineering, University of Alabama at Birmingham, Birmingham, AL 35294, USA.
*Correspondence e-mail: narayana@uab.edu

Extracellular serine protease (Esp) from *Staphylococcus epidermidis* is a glutamyl endopeptidase that inhibits the growth and formation of *S. aureus* biofilms. Previously, crystal structures of the matured and active Esp have been determined. Interestingly, many of the staphylococcal glutamyl endopeptidase zymogens, including V8 from *Staphylococcus aureus* and Esp from *S. epidermidis*, contain unusually long pro-peptide segments; however, their function is not known. With the aim of elucidating the function of these pro-peptide segments, crystal structures of the Esp zymogen (Pro-Esp) and its variants were determined. It was observed that the N-terminus of the Pro-Esp crystal structure is flexible and is not associated with the main body of the enzyme, unlike in the known active Esp structure. In addition, the loops that border the putative substrate-binding pocket of Pro-Esp are flexible and disordered; the structural components that are responsible for enzyme specificity and efficiency in serine proteases are disordered in Pro-Esp. However, the N-terminal locked Pro-Esp variants exhibit a rigid substrate-binding pocket similar to the active Esp structure and regain activity. These structural studies highlight the role of the N-terminus in stabilizing the structural components responsible for the activity and specificity of staphylococcal glutamyl endopeptidases.

1. Introduction

Glutamyl endopeptidases (GSEs) hydrolyze the peptide bonds formed by the α -carboxyl groups of Glu and Asp residues; however, Glu-*X* bonds are cleaved more efficiently than Asp-*X* bonds, where *X* can be any amino acid. Bacterial GSEs belong to the serine protease family and are related to chymotrypsin-like serine proteases. They have been isolated from a number of bacterial species such as *Staphylococcus aureus* (Drapeau *et al.*, 1972), *Staphylococcus epidermidis* (Dubin *et al.*, 2001; Ohara-Nemoto *et al.*, 2002), *Bacillus licheniformis* (Kakudo *et al.*, 1992; Svendsen & Breddam, 1992), *Enterococcus faecalis* (Kawalec *et al.*, 2005) and *Streptomyces griseus* (Suzuki *et al.*, 1994; Svendsen *et al.*, 1991). The V8 protease from *S. aureus* and SprE from *E. faecalis* have been identified as important virulence factors (Calander *et al.*, 2008; Kawalec *et al.*, 2005; Ohara-Nemoto *et al.*, 2002; Qin *et al.*, 2000) along with many other members of this enzyme family. The V8 protease is essential for the activation of the cysteine protease SspB, a virulence factor of *S. aureus* (Massimi *et al.*, 2002). V8 also exhibits 25% sequence identity to toxins A and B (ETA and ETB). It has been implicated in staphylococcal scalded skin syndrome (Bukowski *et al.*, 2010; Dancer *et al.*, 1990; Lee *et al.*, 1987; O'Toole & Foster, 1987). Interestingly, *S. epidermidis* GSE, also known as extracellular

serine protease (Esp), has 69% sequence identity to V8 and has been shown to inhibit biofilm formation and nasal colonization by *S. aureus* (Iwase *et al.*, 2010).

Many crystal structures of mature GSEs are available. The crystal structures of ETA, ETB (Cavarelli *et al.*, 1997) and V8 from *S. aureus* (Prasad *et al.*, 2004), BIEP from *B. intermedius* (Meijers *et al.*, 2004) and Esp from *S. epidermidis* (Chen *et al.*, 2013) reveal that they exhibit a unique substrate-binding pocket; their N-terminus is implicated in the specificity of the enzyme towards negatively charged (Glu and Asp) residues of the substrate.

The activation pathway of chymotrypsin-like serine proteases has been extensively studied, in which the inactive enzyme is generally converted to its mature form by losing the N-terminal pro-peptide segment (Neurath & Walsh, 1976). Interestingly, the pro-peptide segments of bacterial GSEs are unusually long, at more than 60 residues, and could function as inhibitors of the proteolytic domain (Gasarov *et al.*, 2008; Nickerson *et al.*, 2007; Ohara-Nemoto *et al.*, 2008), similar to as seen in human cysteine proteases, where the pro-peptide acts as an intramolecular chaperone (Wiederanders *et al.*, 2003) and as an inhibitor keeping the protease inactive (Khan & James, 1998). Deletion and mutational analyses suggested a similar chaperone-like role for the extra-long pro-peptide of *S. aureus* V8 (Ohara-Nemoto *et al.*, 2008). Meanwhile, the GSE from *S. griseus* has been suggested to be activated autocatalytically by a cleavage between the N-terminal pro-peptide and the mature domain (Rouf *et al.*, 2012; Stennicke *et al.*, 1996; Svendsen *et al.*, 1991); however, the activation of the GSEs from *S. aureus* (Drapeau, 1978; Nickerson *et al.*, 2007; Ono *et al.*, 2008; Rouf *et al.*, 2012), *B. intermedius* (Gasarov *et al.*, 2008; Park *et al.*, 2004; Trachuk *et al.*, 2005) and *Enterococcus* (Kawalec *et al.*, 2005) involves both autocatalytic truncation of the pro-peptide further towards the N-terminus and a final heterocatalytic cleavage of the remaining part of the pro-peptide by other proteases. Modification of the heterocatalytic cleavage site of *B. intermedius* GSE has been shown to affect the production of active GSE by *B. subtilis* cells (Velishaeva *et al.*, 2008), suggesting that the cellular production of active GSE largely relies on the pro-peptide structure and its cleavage sites.

Although Esp from *S. epidermidis* is known to eradicate even pre-existing *S. aureus* biofilms (Iwase *et al.*, 2010), the highly homologous *S. aureus* V8 exhibits no such inhibitory effect on *S. epidermidis* biofilms or colonization. In the light of the high morbidity and mortality associated with multi-drug-resistant strains of *S. aureus* (MRSA; Kleven *et al.*, 2007), there is an urgent need for the identification of a new class of inhibitors that do not elicit bacterial resistance. Understanding the structural correlation of Esp maturation and its functional role in preventing *S. aureus* biofilm formation and growth may provide such a lead.

Our laboratory has previously determined the three-dimensional crystal structure of active Esp by molecular-replacement methods using *S. aureus* V8 (PDB entry 1qy6; Prasad *et al.*, 2004) as a starting model (Chen *et al.*, 2013). Similar to the eukaryotic serine proteases of the chymotrypsin

family, Esp exhibits a β -barrel fold with two domains (each consisting of six antiparallel β -strands) and an α -helix at the C-terminus (Perona *et al.*, 1995; Perona & Craik, 1995; Rühlmann *et al.*, 1973). The essential functional features of the putative catalytic triad residues (Ser235, Asp159 and His117), the oxyanion hole and the S1 pocket (substrate-binding region) are present. However, distinct from the eukaryotic serine proteases, the conserved intra-domain disulfide bonds responsible for structural rigidity are absent in Esp and V8 (Chen *et al.*, 2013; Perona & Craik, 1995; Prasad *et al.*, 2004). Nemoto and coworkers identified the catalytic Ser237 (Ser235 in Esp) and N-terminal Val69 (Val67 in Esp) of V8 as essential for substrate cleavage (Ohara-Nemoto *et al.*, 2008). N-terminal truncation to Ile70 (Ile68 in Esp) renders V8 inactive and mutants with an altered N-terminus, replacing Val69 with conserved substitutions, were also inactive, indicating a strict requirement for the N-terminal Val69 residue (Ono *et al.*, 2010). Moreover, we investigated the mechanism of Esp-mediated *S. aureus* biofilm inhibition and identified autolysin (Atl) as the major target of Esp during biofilm disassembly (Chen *et al.*, 2013). Although the biofilm formed by an *S. aureus* atl mutant was dramatically reduced compared with that of the wild-type strain, the residual biofilm formed by the atl mutant was sensitive to Esp, suggesting that Esp also degrades other biofilm formation-associated proteins such as coagulase (Coa), protein A (SpA) and clumping factor A (ClfA) (Chen *et al.*, 2013).

The disposition of the N-terminus, which associates with the respective S1 pockets more intimately than in other active serine proteases (Prasad *et al.*, 2004; Chen *et al.*, 2013), is identical in V8 and Esp. The amino group of the N-terminal Val67 of Esp (Val69 in V8) is suitably positioned to act as an acceptor of the negative charge of the side chain of the P1 residue Glu. Similarly, the His250 (His252 in V8), Tyr226 (Tyr228 in V8) and Thr230 (Thr232 in V8) residues are highly conserved across GSEs. Modification of His250, which makes hydrogen bonds to the side chains of the conserved Tyr226 (Tyr228 in V8) and Thr230 (Thr232 in V8), significantly affects the catalytic activity (k_{cat} decreases more than 600-fold) but has a low impact on substrate binding (K_m increases approximately fivefold; Demidyuk *et al.*, 2004), suggesting that it may not play a crucial role in holding the negatively charged substrate residue, but it seems to be significant for accurate positioning of the scissile bond with respect to the nucleophile Ser235. Modification of the N-terminal Val69 in V8 has been shown to substantially reduce the hydrolytic efficiency of Glu-carrying substrates, indicating its importance. Furthermore, its α -amino group probably makes a very significant contribution to the recognition of the charged substrate (Nemoto *et al.*, 2008; Ono *et al.*, 2010).

As mentioned earlier, the pro-enzymes of Esp and V8, and all other known bacterial glutamyl endopeptidases (Kawalec *et al.*, 2005; Liu *et al.*, 2016), are equipped with unusually long 68–69-residue N-termini, similar to some cysteine proteases, that need to be clipped for enzyme activation. However, their structural role in the transition from zymogen to active enzyme and the structural changes that the enzyme undergoes

Table 1
Recombinant production information.

Construct	Plasmid	N- and C-terminal tags	Mutations
Full-length Esp, Ala27–Gln282	pET-28b; kanamycin-resistant	His ₆ tag at the N-terminus	NA
Full-length EspS235A, Val67–Gln282	pET-28b; kanamycin-resistant	His ₆ tag at the N-terminus	Ser235 to Ala
EspΔ55, Ile56–Gln282	pMCSG7 ligation-independent cloning; ampicillin-resistant	His ₆ tag at the N-terminus followed by TEV cleavage site	NA
EspΔ55S235A, Ile56–Gln282	pMCSG7 ligation-independent cloning; ampicillin-resistant	His ₆ tag at the N-terminus followed by TEV cleavage site	Ser235 to Ala
EspΔ55S66V, Ile56–Gln282	pMCSG7 ligation-independent cloning; ampicillin-resistant	His ₆ tag at the N-terminus followed by TEV cleavage site	Ser66 to Val
EspΔ55CysM1, Ile56–Gln282	pET-21b; ampicillin-resistant	His ₆ tags at the N- and C-termini with TEV and thrombin cleavage sites, respectively	Val67 to Cys; Thr230 to Cys
EspΔ65CysM1, Ser66–Gln282	pET-21b; ampicillin-resistant	His ₆ tags at the N- and C-termini with TEV and thrombin cleavage sites, respectively	Val67 to Cys; Thr230 to Cys
EspΔ65CysM2, Ser66–Gln282	pET-21b; ampicillin-resistant	His ₆ tags at the N- and C-termini with TEV and thrombin cleavage sites, respectively	Val67 to Cys; Thr230 to Cys
EspΔ58CysM2, Ser59–Gln282	pET-21b; ampicillin-resistant	His ₆ tags at the N- and C-termini with TEV and thrombin cleavage sites, respectively	Val67 to Cys; Asp255 to Cys
EspΔ55CysM2, Ile56–Gln282	pET-21b; ampicillin-resistant	Thrombin cleavage site followed by His ₆ tag at the C-terminus	Val67 to Cys; Asp255 to Cys

owing to the loss of pro-peptide are unclear; the pro-peptide could either be flexible, free-floating as observed in the zymogens of chymotrypsin-like serine proteases, or rigid and intimately associated with the protease body and act as an inhibitor, as seen in cysteine protease zymogens. In this paper, we determine crystal structures of Pro-Esp and its variants to localize the pro-peptide in relation to the main body of the enzyme and to understand the structure–function relationships of such an extended N-terminus and specifically the residues responsible for Pro-Esp inactivity and activation.

2. Materials and methods

2.1. Cloning and mutations

Full-length Pro-Esp was cloned in the pET-28b vector, as described previously (Chen *et al.*, 2013). Other constructs of Pro-Esp were cloned in pET-28b, pMCSG7 and pET-21b plasmids using XhoI and NdeI (New England BioLabs) restriction sites with a His₆ tag at the N-terminus and/or the C-terminus. Mutations were made using the Change-IT Multiple Mutation Site-Directed Mutagenesis Kit (USB Affymetrix; Chen & Ruffner, 1998). The clones obtained and the mutations performed were confirmed by DNA sequencing (Table 1).

2.2. Expression and purification

The respective plasmids were transformed into *Escherichia coli* BL21 (DE3) cells for expression. 1 l Luria–Bertani (LB) broth supplemented with 100 mg ml^{−1} ampicillin or 50 mg ml^{−1} kanamycin (depending on the plasmid) was inoculated with an overnight culture (10 ml) of stationary-phase *E. coli*. The cells were grown at 37°C until the culture reached an OD_{600 nm} of ~0.5, at which point the temperature was reduced to 30°C. When the OD_{600 nm} reached ~0.7, protein expression was induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and the culture was incubated overnight at 30°C. Harvesting of the cells was carried out by

centrifugation at 3210g (Beckman Allegra 6R Centrifuge) for 30 min at 4°C; the cells were resuspended in lysis buffer consisting of 50 mM Tris pH 8.0, 250 mM NaCl, 10%(v/v) glycerol and an EDTA-free protease-inhibitor cocktail tablet (Roche). The cells were lysed by sonication on ice and the soluble supernatant was collected by centrifugation at 48 384g (Beckman Avanti J-25 centrifuge) for 1 h at 4°C. The clear lysate was allowed to pass through a pre-packed 5 ml HisTrap FF column (GE Healthcare) pre-equilibrated with lysis buffer at a flow rate of 1 ml min^{−1} to enable binding of the recombinant His-tagged proteins to the beads. The protein-bound column was washed with ten bed volumes of buffer A (50 mM Tris pH 8.0, 50 mM NaCl), followed by a wash with ten bed volumes of buffer consisting of 50 mM pH 8.0, 50 mM NaCl, 37.5 mM imidazole. The bound protein was eluted with a linear gradient of 15–100% buffer B and the fractions containing the recombinant protein were checked for purity and molecular weight on a 12% SDS–PAGE gel (Laemmli, 1970). These samples were further purified by gel-filtration chromatography (HiPrep 16/60 Sephacryl S-100 HR column; GE Healthcare) with 50 mM Tris pH 8.0, 150 mM NaCl, 10% glycerol.

2.3. Enzyme activity

The activity of the purified Pro-Esp variant recombinants was tested using azocasein, a nonspecific substrate (Rice *et al.*, 2001). Proteins (40 µg per reaction) were incubated at 37°C for 1 h with 1%(w/v) azocasein (Sigma) in a total volume of 400 µl. Following incubation, 400 µl 5% trichloroacetic acid (TCA) was added to quench the reaction. Upon centrifugation at 14 000g for 10 min, the soluble material was used to check the activity by measuring the absorbance at 440 nm using a spectrophotometer.

The enzyme activity of purified Pro-Esp and its variant recombinants (0.2 µg per reaction) was measured using 1 mM Z-Leu-Leu-Glu-AMC (where Z is a carboxyl/Boc protecting group and AMC is 7-amino-4-methylcoumarin, a detectable fluorescent group; Enzo Life Sciences) in 50 mM Tris–HCl pH

Table 2

Crystallization conditions and data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	EspS235A†	EspΔ55S235A-pro	EspΔ55S66V-pro	EspΔ65CysM2†	EspΔ58CysM2†
Crystallization conditions					
Composition of reservoir solution	20% PEG 8000, 0.1 M HEPES pH 7.5	0.1 M magnesium chloride, 0.1 M sodium acetate pH 4.6, 30% PEG 4000	20% PEG 3350, 0.225 M ammonium nitrate	25% PEG MME 5000, 0.2 M lithium sulfate, 0.1 M Tris pH 8.5	1 M lithium chloride, 0.1 M Tris pH 8.5, 20% PEG 6000
Cryoprotectant	15% glycerol	20% glycerol	35% ethylene glycol	15% glycerol	20% PEG 400
Data collection					
Wavelength (Å)	1.0	1.0	1.54	1.0	1.0
Resolution range (Å)	34.6–1.200 (1.243–1.200)	30.4–1.850 (1.916–1.850)	38.4–2.199 (2.278–2.199)	36.8–2.066 (2.140–2.066)	36.2–2.080 (2.154–2.080)
Space group	<i>P</i> 12 ₁ 1	<i>P</i> 4 ₁ 2 ₁ 2	<i>P</i> 12 ₁ 1	<i>P</i> 6 ₁ 22	<i>P</i> 6 ₁ 22
<i>a</i> , <i>b</i> , <i>c</i> (Å)	39.6, 61.0, 42.5	69.1, 69.1, 127.3	42.6, 61.1, 75.5	42.5, 42.5, 373.4	42.0, 42.0, 378.4
α , β , γ (°)	90.0, 98.7, 90.0	90.0, 90.0, 90.0	90.0, 94.7, 90.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0
No. of molecules in the asymmetric unit	1	1	2	1	1
Total reflections	119842 (10950)	47575 (4664)	28089 (2054)	193282 (4750)	25524 (1948)
Unique reflections	60633 (5576)	25463 (2522)	18892 (1779)	13121 (1075)	12812 (1016)
Multiplicity	2.0 (2.0)	1.9 (1.8)	1.5 (1.2)	14.7 (4.4)	2.0 (1.9)
Completeness (%)	97 (89)	94 (95)	95 (91)	97 (82)	97 (81)
Mean <i>I</i> / σ (<i>I</i>)	30.8 (13.8)	8.8 (1.2)	9.7 (4.5)	49.0 (17.2)	39.7 (13.4)
Wilson <i>B</i> factor (Å ²)	9.8	23.9	11.8	16.9	19.3
<i>R</i> _{merge} (%)	1.5 (3.8)	6.2 (61.8)	4.1 (8.8)	4.2 (5.6)	1.0 (3.8)
<i>R</i> _{meas} (%)	2.2 (5.4)	8.8 (87.4)	5.8 (12.5)	4.4 (6.2)	1.4 (5.4)
<i>CC</i> _{1/2}	99.8 (99.5)	99.7 (43.6)	99.4 (96.6)	99.8 (99.7)	99.9 (99.2)
<i>CC</i> [*]	99.9 (99.9)	99.9 (77.9)	99.8 (99.1)	100.0 (99.9)	100.0 (99.8)
Refinement					
Reflections used in refinement	60628 (5576)	25460 (2522)	18887 (1779)	13114 (1075)	12810 (1014)
Reflections used for <i>R</i> _{free}	1799 (166)	1999 (199)	980 (118)	1312 (108)	1281 (101)
<i>R</i> _{work} (%)	15.9 (17.0)	19.3 (28.7)	17.0 (17.5)	14.7 (16.3)	16.0 (19.4)
<i>R</i> _{free} (%)	16.4 (18.2)	22.2 (31.5)	23.2 (25.7)	18.6 (21.7)	21.5 (24.8)
<i>CC</i> (work)	0.95 (0.93)	0.96 (0.72)	0.95 (0.92)	0.96 (0.93)	0.95 (0.91)
<i>CC</i> (free)	0.94 (0.93)	0.95 (0.73)	0.91 (0.85)	0.94 (0.90)	0.92 (0.87)
No. of non-H atoms					
Protein	1646	1620	3145	1670	1704
Water	299	190	275	207	186
Total	1945	1810	3420	1877	1890
R.m.s. deviations					
Bonds (Å)	0.004	0.007	0.007	0.009	0.007
Angles (°)	0.83	0.82	0.79	0.86	0.82
Ramachandran plot					
Most favored (%)	96.2	96.6	96.4	95.8	96
Allowed (%)	3.8	3.4	3.6	4.2	3.7
Outliers (%)	0	0	0	0	0
PDB code	6pym	6q24	6q12	6tya	6u1b

† Friedel mates were averaged when calculating reflection statistics.

8.0 containing 5 mM EDTA in a total volume of 200 µl. EDTA was added to suppress the activity of thermolysin and other nonspecific metalloproteases (Fontana, 1988). The reaction mixture was incubated at 37°C for 1 h before measuring the fluorescence with excitation at 370 nm and emission at 460 nm using a FLUOstar OPTIMA (BMG Labtech) in a 384-well hard-shell black plate.

2.4. Protein crystallization and X-ray diffraction data collection

Purified Pro-Esp and its variants were concentrated using an Amicon ultracentrifugation system and initial crystallization conditions were screened by the sitting-drop vapor-diffusion method at 20°C using a Crystal Phoenix robot (Art Robbins Instruments). Commercially available crystallization screens such as The PEGs and PEGs II Suites (Qiagen), JCSG Core I and II Suites (Qiagen), Crystal Screen HT (Hampton

Research) and Wizard Classic 1 and 2 (Rigaku Reagents) were used for screening. The crystallization conditions were optimized using the hanging-drop vapor-diffusion method at 22°C and good-quality single crystals were produced after optimization. Native diffraction data were collected at 100 K using an appropriate cryoprotectant either in-house or on the SER-CAT 22-ID beamline at the Advanced Photon Source. Optimized crystallization conditions, the cryoprotectant used for data collection at 100 K and the diffraction data statistics for the Esp variants are summarized in Table 2.

2.5. Structure determination

Data indexing, integration and scaling were performed using *XDS* (Kabsch, 2010) and *iMosflm* (Battye *et al.*, 2011) and the crystal structures were solved by the molecular-replacement technique (Rossmann, 1990) using the mature Esp structure (PDB entry 4jcn; Chen *et al.*, 2013) as the

starting model. Model building, refinement and validation were performed with the *CCP4* suite (Collaborative Computational Project, Number 4, 1994; Winn *et al.*, 2011; Vagin *et al.*, 2004; Matthews, 1968), *Phenix* (Liebschner *et al.*, 2019) and *Coot* (Emsley & Cowtan, 2004; Emsley *et al.*, 2010).

2.6. N-terminal sequencing

Protein sequencing was performed using an ABI 494 protein sequencer. Samples were transferred onto a PVDF membrane and shipped to Tufts University core facility for sequencing.

2.7. Sequence and structure comparison

The primary sequences of the Pro-Esp constructs were compared using *MUSCLE* (Edgar, 2004), an online multiple sequence-alignment tool from The European Bioinformatics Institute (EMBL-EBI; McWilliam *et al.*, 2013). Crystal structure comparison was performed using *Coot* (Emsley *et al.*, 2010), *PyMOL* (v.1.8; Schrödinger; <http://www.pymol.org>) and *UCSF Chimera* (Pettersen *et al.*, 2004). Figures were generated using *PyMOL*.

2.8. Docking studies

Molecular-docking studies were performed using an open-source program called *AutoDock Vina* (Trott & Olson, 2010). The high-resolution crystal structure of EspS235A was used as the minimized receptor structure and the peptide Z-Leu-Leu-Glu-AMC was docked into it. The structures were optimized using *AutoDock* parameters such as atom types, torsion modes and partial charges, and the models were saved in PDBQT format. A grid of dimensions $62 \times 34 \times 46$ Å was generated to identify *xyz* coordinates ($x = 19.331$, $y = 19.55$, $z = 1.444$) around the substrate-binding region. To find the most favorable interactions, the Lamarckian genetic algorithm (LGA) was selected with default parameters in *AutoDock Vina*. The best pose was selected depending on the highest binding energy criterion.

3. Results

3.1. Primary sequences, data collection and refinement statistics of Esp variants

The primary sequences of the Pro-Esp constructs resulting from cloning and expression were compared using *T-Coffee* (Notredame *et al.*, 2000) and *BoxShade* from ExPASy (Fig. 1), and the primary structures of the purified proteins are depicted in Fig. 2 with the thermolysin cleavage site, N-terminal pro-peptide and mutations labeled. Purified full-length Pro-Esp and full-length Pro-EspS235A were treated with thermolysin to cleave between Ser66 and Val67 in order to generate mature Esp and mature EspS235A, respectively. Before setting up the crystallization screens, the N- and C-terminal His₆ tags were removed where applicable using TEV protease and thrombin, respectively. Protein purity was checked on an SDS-PAGE gel and the data-collection and

refinement statistics for the protein crystals are given in Table 2.

3.2. Structural comparison of Esp with Pro-Esp

Mature EspS235A was crystallized and its three-dimensional structure was determined by molecular replacement. Except for the Ala235 residue, the mature EspS235A structure was identical to the three-dimensional structure of mature Esp previously determined in our laboratory (PDB entry 4jcn; Fig. 3c; Chen *et al.*, 2013).

It has previously been shown that an undecapeptide of the endogenous pro-peptide (Ile56–Ser66) mimicked the native pro-peptide (66 residues) in suppressing the proteolytic activity completely and supporting efficient cleavage at the Ser66–Val67 bond by thermolysin (Ohara-Nemoto *et al.*, 2008). Firstly, we tried to crystallize recombinant EspΔ55 (Ile56–Gln282), but it consistently underwent autoproteolysis to produce mature Esp crystals. To prevent the autoproteolysis, we cloned two variants. In one variant we mutated the active site Ser235 to Ala (EspΔ55S235A) to prevent autocleavage, and in the other we mutated Ser66 to Val (EspΔ55S66V) to block the cleavage of the Ser66–Val67 bond (Figs. 1 and 2). We were able to determine crystal structures of both EspΔ55S66V and EspΔ55S235A successfully. The EspΔ55S235A recombinant crystallized in space group $P4_12_12$ with one molecule in the asymmetric unit, while the crystals of the EspΔ55S66V recombinant belonged to space group $P2_1$, with two molecules in the asymmetric unit related by a twofold noncrystallographic axis. The root-mean-square deviation (r.m.s.d.) between EspΔ55S235A and EspΔ55S66V was 0.34–0.36 Å, with some interesting structural differences between the two.

The crystal structure of EspΔ55S66V, which is referred to as Pro-Esp from here on, has two molecules in the asymmetric unit (*A* and *B*). Molecules *A* and *B* are almost identical (r.m.s.d. of 0.23 Å) and exhibit significant differences from the mature Esp structure (PDB entry 4jcn). The substrate-binding sites of both Pro-Esp molecules in the asymmetric unit (*A* and *B*) were solvent-exposed. More importantly, no electron density was observed for residues Ser63–Asn72 in both molecules *A* and *B*, and thus the Val67 residue critical for substrate recognition (Nemoto *et al.*, 2008; Ono *et al.*, 2010) was not associated with the S1 pocket as seen in the mature Esp structure. The electron density for the loop 1 residues Ser229–Gly233 in molecule *A* and Leu228–Thr230 in molecule *B* and the loop 2 residues Val254–Lys257 in both molecules *A* and *B* was weak, indicating flexibility (Fig. 4b). Significantly, the disorder in loop 1 of molecules *A* and *B* also extended to the respective oxyanion-hole residues (Val231–Ser235); a disordered oxyanion-hole conformation is deleterious to serine protease activity and efficiency (Perona & Craik, 1995; Fig. 4b).

The EspΔ55S235A recombinant crystallized with one molecule in the asymmetric unit and exhibited no enzyme activity in an azocasein assay, similar to Pro-Esp, and the electron density was also missing for its N-terminal residues Ser63–Asn72. Unlike the Pro-Esp structure, loop 1 and loop 2

of EspΔ55S235A had clear, well defined electron density; loop 2 was almost identical to the mature Esp structure. However, loop 1, which contained the conserved 'oxanion-hole' region, moved inwards compared with the mature Esp structure. This

significant movement was owing to crystal-packing contacts from a neighboring symmetry-related molecule.

Interestingly, the loop 2 residue region of the neighboring symmetry-related molecule is seen to associate with the S1

	30	40	50	60	70	80
full length Esp	AKTDTESHNHSSLGTENKNVLDINSSSHNIKPSQNKSYPSVILPNNNRHQIFNTTQGHYD					
full length EspS235A	AKTDTESHNHSSLGTENKNVLDINSSSHNIKPSQNKSYPSVILPNNNRHQIFNTTQGHYD					
mature Esp	VILPNNNRHQIFNTTQGHYD					
EspS235A	VILPNNNRHQIFNTTQGHYD					
EspΔ55S66V	IKPSQNKSYPVVILPNNNRHQIFNTTQGHYD					
EspΔ55S235A	IKPSQNKSYPSVILPNNNRHQIFNTTQGHYD					
EspΔ65CysM1	SCILPNNNRHQIFNTTQGHYD					
EspΔ65CysM2	SCILPNNNRHQIFNTTQGHYD					
EspΔ58CysM2	SQNKSYPSVILPNNNRHQIFNTTQGHYD					
EspΔ55CysM2	IKPSQNKSYPSVILPNNNRHQIFNTTQGHYD					
	90	100	110	120	130	140
full length Esp	AVSFIYIPIIDGGYMSGSGVVGGENEILTNNKHVVNGAKGNPRNISVHPSAKNENDYPNGKF					
full length EspS235A	AVSFIYIPIIDGGYMSGSGVVGGENEILTNNKHVVNGAKGNPRNISVHPSAKNENDYPNGKF					
mature Esp	AVSFIYIPIIDGGYMSGSGVVGGENEILTNNKHVVNGAKGNPRNISVHPSAKNENDYPNGKF					
EspS235A	AVSFIYIPIIDGGYMSGSGVVGGENEILTNNKHVVNGAKGNPRNISVHPSAKNENDYPNGKF					
EspΔ55S66V	AVSFIYIPIIDGGYMSGSGVVGGENEILTNNKHVVNGAKGNPRNISVHPSAKNENDYPNGKF					
EspΔ55S235A	AVSFIYIPIIDGGYMSGSGVVGGENEILTNNKHVVNGAKGNPRNISVHPSAKNENDYPNGKF					
EspΔ65CysM1	AVSFIYIPIIDGGYMSGSGVVGGENEILTNNKHVVNGAKGNPRNISVHPSAKNENDYPNGKF					
EspΔ65CysM2	AVSFIYIPIIDGGYMSGSGVVGGENEILTNNKHVVNGAKGNPRNISVHPSAKNENDYPNGKF					
EspΔ58CysM2	AVSFIYIPIIDGGYMSGSGVVGGENEILTNNKHVVNGAKGNPRNISVHPSAKNENDYPNGKF					
EspΔ55CysM2	AVSFIYIPIIDGGYMSGSGVVGGENEILTNNKHVVNGAKGNPRNISVHPSAKNENDYPNGKF					
	150	160	170	180	190	200
full length Esp	VGQEIIPYPGNSDLAILRVSPNEHNQHQIGQVVKPATISSNTDTRINENITVTGYPGDKPL					
full length EspS235A	VGQEIIPYPGNSDLAILRVSPNEHNQHQIGQVVKPATISSNTDTRINENITVTGYPGDKPL					
mature Esp	VGQEIIPYPGNSDLAILRVSPNEHNQHQIGQVVKPATISSNTDTRINENITVTGYPGDKPL					
EspS235A	VGQEIIPYPGNSDLAILRVSPNEHNQHQIGQVVKPATISSNTDTRINENITVTGYPGDKPL					
EspΔ55S66V	VGQEIIPYPGNSDLAILRVSPNEHNQHQIGQVVKPATISSNTDTRINENITVTGYPGDKPL					
EspΔ55S235A	VGQEIIPYPGNSDLAILRVSPNEHNQHQIGQVVKPATISSNTDTRINENITVTGYPGDKPL					
EspΔ65CysM1	VGQEIIPYPGNSDLAILRVSPNEHNQHQIGQVVKPATISSNTDTRINENITVTGYPGDKPL					
EspΔ65CysM2	VGQEIIPYPGNSDLAILRVSPNEHNQHQIGQVVKPATISSNTDTRINENITVTGYPGDKPL					
EspΔ58CysM2	VGQEIIPYPGNSDLAILRVSPNEHNQHQIGQVVKPATISSNTDTRINENITVTGYPGDKPL					
EspΔ55CysM2	VGQEIIPYPGNSDLAILRVSPNEHNQHQIGQVVKPATISSNTDTRINENITVTGYPGDKPL					
	210	220	230	240	250	260
full length Esp	ATMWESVGKVYIIGGEELRYDLS TVGGNSGSPVFNGKNQVIGIHYGGVDNKNYSSVYIND					
full length EspS235A	ATMWESVGKVYIIGGEELRYDLS TVGGNSGSPVFNGKNQVIGIHYGGVDNKNYSSVYIND					
mature Esp	ATMWESVGKVYIIGGEELRYDLS TVGGNSGSPVFNGKNQVIGIHYGGVDNKNYSSVYIND					
EspS235A	ATMWESVGKVYIIGGEELRYDLS TVGGNSGSPVFNGKNQVIGIHYGGVDNKNYSSVYIND					
EspΔ55S66V	ATMWESVGKVYIIGGEELRYDLS TVGGNSGSPVFNGKNQVIGIHYGGVDNKNYSSVYIND					
EspΔ55S235A	ATMWESVGKVYIIGGEELRYDLS TVGGNSGSPVFNGKNQVIGIHYGGVDNKNYSSVYIND					
EspΔ65CysM1	ATMWESVGKVYIIGGEELRYDLS TVGGNSGSPVFNGKNQVIGIHYGGVDNKNYSSVYIND					
EspΔ65CysM2	ATMWESVGKVYIIGGEELRYDLS TVGGNSGSPVFNGKNQVIGIHYGGVDNKNYSSVYIND					
EspΔ58CysM2	ATMWESVGKVYIIGGEELRYDLS TVGGNSGSPVFNGKNQVIGIHYGGVDNKNYSSVYIND					
EspΔ55CysM2	ATMWESVGKVYIIGGEELRYDLS TVGGNSGSPVFNGKNQVIGIHYGGVDNKNYSSVYIND					
	270	280				
full length Esp	FVQQFLRNNIPDINIQ					
full length EspS235A	FVQQFLRNNIPDINIQ					
mature Esp	FVQQFLRNNIPDINIQ					
EspS235A	FVQQFLRNNIPDINIQ					
EspΔ55S66V	FVQQFLRNNIPDINIQ					
EspΔ55S235A	FVQQFLRNNIPDINIQ					
EspΔ65CysM1	FVQQFLRNNIPDINIQ					
EspΔ65CysM2	FVQQFLRNNIPDINIQ					
EspΔ58CysM2	FVQQFLRNNIPDINIQ					
EspΔ55CysM2	FVQQFLRNNIPDINIQ					

Figure 1

Primary-sequence comparison of recombinant Esp variants. Sequence similarity is highlighted in red and the mutations S66V, V67C, T230C, S235A and D255C in their respective constructs are indicated in boxes. The missing N-terminal residues are indicated by dotted lines.

pocket of the Esp Δ 55S235A reference molecule. The side chain of Asp255 in the intruding loop is suitably positioned to

Figure 2
Schematic representation of the structures of recombinant Esp variants. The thermolysin cleavage site is indicated with an arrow and the N-terminal pro-peptide residues are labeled. Mutated residues are indicated with red lines and labeled.

interact with the conserved His250 of the S1 pocket, which has been implicated in positioning the P1 residue Glu in the mature Esp–substrate complex (Demidyuk *et al.*, 2004). Also, the Asp255–Asn256 peptide bond of the intruding loop is positioned such that the carbonyl atom of Asp255 is at a suitable distance (~ 3 Å) for nucleophilic attack by the catalytic Ser235. This serendipitous observation may provide a partial glimpse of the substrate-bound active-site structure of Esp (Fig. 5).

Thus, both the Esp Δ 55S235A and Esp Δ 55S66V (Pro-Esp) structures revealed a flexible N-terminus and a distorted loop 1 and loop 2 in the S1 pocket, similar to other chymotrypsin-like serine protease zymogen structures.

(a)

Figure 3
(a) EspS235A crystals. (b) Crystal structure overlay of EspS235A (blue) and mature Esp (PDB entry 4jcn; green). His117, Asp159 and Ala235 are represented as sticks in magenta. Loops 1 and 2 are labeled.

3.3. Crystal structures of locked N-terminal Pro-Esp variants

To check the significance of such a uniquely positioned N-terminus, associating with and covering the bottom of the S1 pocket, as seen in the structures of mature GSE enzymes, we created cysteine mutants of Pro-Esp. We expected the enzyme activity to be retained when the N-terminus is locked in a position similar to mature Esp but with extra N-terminal residues. Firstly, we mutated Val67 and Thr230 (loop 2) to cysteine (Esp Δ 65CysM1; Figs. 1 and 2) and crystallized the recombinant. However, the crystal structure of the Esp Δ 65CysM1 recombinant revealed no disulfide bond between Cys67 and Cys230, and there was no electron density for residues Ser66–Asn72 of the N-terminus, similar to the Pro-Esp crystal structure.

Although there was no interaction between Asp255 on loop 2 and the N-terminal Val67, their C α atoms were in close proximity to form a disulfide bond when mutated to cysteines (Esp Δ 65CysM2; Figs. 1 and 2). The crystal structure of Esp Δ 65CysM2 containing just one residue from the pro-segment (Ser66) revealed a disulfide bond between the side-chain –SH groups of residues Cys67 and Cys255. Furthermore,

loops 1 and 2 were stabilized and the oxyanion hole adopted a functionally active and open conformation.

The crystal structure of Esp Δ 58CysM2, similar to that of Esp Δ 65CysM2, had a disulfide bond between the side-chain –SH groups of Cys67 and Cys255. It also showed clearly defined electron densities for the residues involved in the formation of loops 1 and 2. The oxyanion hole in this variant also adopted an open and functionally active conformation similar to that in mature Esp and Esp Δ 66CysM2. Additionally, we observed that the six residues in the pro-peptide segment, Asn61–Ser66, were positioned on top of the S1 pocket, potentially blocking its access to the substrate.

4. Discussion

4.1. A structural perspective on Pro-Esp activation

It has previously been reported that residues Met1–Ala27 of the native Esp form a pre-segment containing a signal sequence and that Lys28–Ser66 is a pro-segment, suggesting that Esp is secreted as a 32 kDa (Lys28–Gln282) pro-enzyme (Ohara-Nemoto *et al.*, 2008). Therefore, we cloned the recombinant pro-enzyme Ala27–Gln282 (Fig. 3) but, as detected by N-terminal sequencing (Table 3), the purified pro-enzyme converted to a 29 kDa intermediate (Ile49–Gln282). This intermediate pro-enzyme is referred to as full-length Esp in the subsequent discussion. Additionally, Table 3 details the regions with undefined electron density at the N-terminus of the respective crystal structures.

We were not able to crystallize full-length Esp, perhaps owing to the unusually long and flexible pro-peptide. According to Nemoto *et al.* (2008), the undecapeptide of the endogenous pro-peptide was adequate to completely suppress

Figure 4

(a) Pro-Esp crystals. (b) Crystal structure overlay of Pro-Esp (red) and mature Esp (green). The catalytic triad and the N-terminal residues Asn73 in Pro-Esp and Val67 in mature Esp are represented as sticks in magenta. The N-terminus of Pro-Esp is repositioned, destabilizing loops 1 and 2. The oxyanion hole is in a nonfunctional conformation. (b) is enlarged to highlight the residues on the N-terminus and loops 1 and 2.

Figure 5

The crystal structure of Esp Δ 55S235A (cyan) is superimposed (r.m.s.d. of 0.41 Å) on that of mature Esp (PDB entry 4jcn; orange). Val67, His117, Asp159 and Ser235 of mature Esp are labeled and represented as sticks in red. In Esp Δ 55S235A, Asn73 and His250 (represented as sticks in red) are labeled. Loop 2 of the symmetry-related molecule (magenta) of Esp Δ 55S235A is indicated with an arrow, and Asp255 and Asn256 are labeled and represented as sticks in yellow. The inward movement of loop 1 and the closed conformation of the oxyanion hole are shown as an arrow and a dotted ellipse, respectively. The contacts between residues Asp255–His250 (2.82 Å) and Asn256–Ser235 (2.52 Å) are represented as dotted lines.

Table 3
N-terminal sequencing and definition from electron density.

Esp variant	N-terminal sequencing		Crystal	N-terminus in the crystal structure (electron density)
	Sample			
Full-length Esp	I ⁴⁹ NSSS ⁵³ , missing residues Ala27–Asp48		NA	NA
Esp mature	NA		NA	V ⁶⁷ ILPN ⁷¹
EspS235A	NA		NA	V ⁶⁷ ILPN ⁷¹
EspΔ55S235A	NA		NA	N ⁷³ RHQI ⁷⁷ , undefined density for residues Ile56–Asn72
EspΔ55S66V	S ⁶³ YPVV ⁶⁷ , missing residues Ile56–Lys62		NA	N ⁷³ RHQI ⁷⁷ , undefined density for residues Ser63–Asn72
EspΔ55CysM1	SI ⁵⁶ KPS ⁵⁹		NA	NA
EspΔ65CysM1	S ⁶⁶ CILP ⁷⁰		S ⁶⁶ CILP ⁷⁰	N ⁷³ RHQI ⁷⁷ , undefined density for residues Ser66–Asn72
EspΔ65CysM2	S ⁶⁶ CILP ⁷⁰		NA	S ⁶⁶ CILP ⁷⁰
EspΔ58CysM2	N ⁶¹ KSYP ⁶⁵ , missing residues Ser59 and Glu60		N ⁶¹ KSYP ⁶⁵	N ⁶¹ KSYP ⁶⁵
EspΔ55CysM2	MI ⁵⁶ KPS ⁵⁹		MI ⁵⁶ KPS ⁵⁹	P ⁶⁵ SCIL ⁶⁹ , undefined density for residues Ile56–Tyr64

the proteolytic activity and to support efficient cleavage at the Ser66–Val67 bond by thermolysin (Ohara-Nemoto *et al.*, 2008). Interestingly, full-length Esp displayed residual activity in the azocasein assay which was about 35-fold lower than that of mature Esp (Fig. 6*a*). This was similar to the zymogens of the staphylococcal SplB protease (Pustelny *et al.*, 2014) and bovine trypsin (Fehlhammer *et al.*, 1977), in which the spatial arrangement of the catalytic residues (Ser214, Asp102, His57 and Ser195; chymotrypsin numbering) in the zymogens was almost similar, but with slightly distorted or inadequately formed S1 pockets compared with the mature enzymes. Such an arrangement in the zymogen could also explain the complete absence of activity against the specific substrate (Fig. 6*b*) for the full-length Esp recombinant. The EspS235A recombinant showed no activity against azocasein and specific substrate as expected, but this raised a pertinent question. Does the long pro-peptide in full-length Esp block the active site by association as in cathepsins (Fox *et al.*, 1992), but is not tight enough and ‘leaks’ residual activity, or is it as in trypsinogen with a flexible N-terminus and a disordered S1 pocket, hence the weak enzyme activity?

As expected, the purified EspΔ55S235A and EspΔ55S66V recombinants were inactive (Figs. 7*a* and 7*b*) despite the loss of seven residues (Ile56–Lys62) in EspΔ55S66V, as identified from N-terminal sequencing (Table 3). The crystal structures of both variants showed density starting from Asn73 of the N-terminus, suggesting a flexible nature for the pro-segment. The stabilization of the N-terminus beyond Asn73 was mainly owing to the presence of Arg74 and its salt bridge with Asp230. Interestingly, in the mature Esp structure and also in the EspΔ55S235A structure loops 1 and 2 had clear density, and Arg74, in addition to Asp230, also interacted with the Asn234 side chain through a hydrogen-bonding network mediated by a conserved water molecule. This Arg74–Asn234 interaction was reminiscent of and was spatially similar to the interaction between the N-terminal Ile NH₂ positive charge and Asp194 (chymotrypsin numbering), as seen in all activated serine proteases, an interaction that is responsible for stabilization of the S1 pocket and for the critical ‘oxanion-hole’ formation during conversion from the zymogen to the active enzyme (Volanakis & Narayana, 1996). However, the N-terminal positive charge in GSEs, a product of pro-peptide loss during the transformation from the zymogen to the active

enzyme, was directly involved in the stable association with the S1 pocket but not through interaction with Asn234 (Asp194; chymotrypsin numbering). It is interesting to note that owing to Arg74–Asn234 interactions, loop 1 in EspΔ55S235A had clear density. However, in the absence of anchoring by the N-terminal Val67, this particular loop, which was highly conserved spatially in all serine proteases, had shifted inwards to accommodate the crystallographic contacts and substrate-like peptide (Fig. 5) intrusion into the S1 pocket.

4.2. Elucidating the significance of the N-terminal Val67 residue in the activity

The position of the N-terminus, as observed in the Pro-Esp structure, begged the questions: if we secure the N-terminus in

Figure 6
Protease activities (mean ± SD, n = 3) of Esp variants before and after incubation with thermolysin overnight at room temperature using (a) the nonspecific substrate azocasein and (b) the specific substrate Z-LLE-AMC. Samples are labeled as follows: 1, full-length Esp; 2, full-length Esp + thermolysin; 3, EspS235A; 4, EspΔ55; 5, EspΔ55 + thermolysin; 6, EspΔ55S235A; 7, EspΔ55S66V (Pro-Esp); 8, EspΔ55CysM1; 9, EspΔ65CysM2; 10, EspΔ58CysM2; 11, EspΔ55CysM2. Full-length Esp and EspΔ55 are activated after cleavage at the thermolysin site.

its preferred position, will it stabilize loops 1 and 2 and can the enzyme activity be regained? One way to address these questions was by securing the N-terminus with disulfide bonds. Our initial attempts to lock the N-terminus by mutating Val67 to Cys and Thr230 of loop 1 to Cys (Esp Δ 65CysM1) failed. We suspect that the side chains of Cys67 and Cys230 were not optimally oriented to form a disulfide bond, which could reinforce the significance of the spatial orientation and positioning of Val67, as suggested by Ono *et al.* (2010); any replacement of the Val69 residue of V8 (Val67 in Esp) resulted in a loss of the enzyme activity. The crystal structure of Esp Δ 65CysM1 was similar to that of Pro-Esp, in which electron density for Ser66–Asn72 was missing despite the presence of these residues before and after crystallization (Table 3).

Next, we locked the N-terminus by replacing Val67 and Asp255 (loop 2) with Cys residues. This purified recombinant Pro-Esp variant with eight pro-peptide residues Ser59–Ser66 (Esp Δ 58CysM2) had no enzymatic activity (Figs. 6*a* and 6*b*). Even though N-terminal sequencing of the Esp Δ 58CysM2

recombinant revealed a loss of the Ser59 and Gln60 residues (Table 3), the Esp Δ 58CysM2 crystal structure revealed an S1 pocket and catalytic apparatus similar to those of mature Esp with the six-residue pro-segment Asn61–Ser66 blocking the S1 pocket (Fig. 7*b*). Interestingly, the Esp Δ 65CysM2 recombinant regained partial activity in the azocasein assay, which was tenfold higher than that of full-length Esp and only 2.5-fold lower than that of mature Esp (Fig. 6*a*). The Esp Δ 65CysM2 crystal structure, in contrast to that of Esp Δ 65CysM1, displayed a disulfide bond between Cys67 and Cys255 and stabilized loops 1 and 2 and the overall structure (Fig. 7*a*). It also resembled that of mature Esp. However, Esp Δ 65CysM2 retained one residue from the pro-segment (Ser66) owing to cloning constraints. Upon inserting the TEV cleavage site (ENLYFQS) and treating the Esp Δ 65CysM2 protein with TEV protease, the bond between the Gln and Ser residues was cleaved, leaving the serine from the TEV cleavage site at the N-terminus. Serendipitously, this matched with the Ser66 residue on the pro-peptide of Esp. The reduced activity of

Figure 7

(*a*) Esp Δ 58CysM2 (cyan) is superimposed onto the Pro-Esp structure (red). (*b*) Esp Δ 65CysM2 (yellow) is superimposed onto Esp Δ 58CysM2 (cyan). The disulfide bonds between Cys67 and Cys255 are represented as sticks in yellow in (*a*) and magenta in (*b*). In both of the CysM2 structures the N-terminus is locked stabilizing loops 1 and 2, and the oxyanion hole is in the functional conformation. The pro-peptide extension blocks the S1 pocket in Esp Δ 58CysM2. (*a*) and (*b*) are enlarged to highlight the residues at the N-terminus and loops 1 and 2. (*c*) Crystals of the N-terminal locked Esp variant.

recombinant Esp Δ 65CysM2 could be explained with the help of its crystal structure, in which the terminal Ser66 of the locked N-terminus was seen to partially block the S1 pocket. The side-chain hydroxyl group of Ser66 interacted with the catalytic Ser235 through a network of water-mediated hydrogen bonds. Owing to the high structural similarity between Esp Δ 65CysM2 and mature Esp, we can suggest that a Pro-Esp cysteine mutant without the extra Ser66 residue would regain full activity. The crystal structures of all of the Esp variants discussed so far were superimposed onto that of EspS235A for comparison (Fig. 8).

Figure 8
Superimposing the crystal structures of all of the Esp variants on EspS235A (blue) enlarged to highlight the residues at the N-terminus and loops 1 and 2. Pro-Esp (Esp Δ 55S66V) is in red (r.m.s.d. of 0.2 Å), Esp Δ 65CysM2 is in yellow (r.m.s.d. of 0.15 Å) and Esp Δ 58CysM2 is in cyan (r.m.s.d. of 0.2 Å). Disulfide bonds and catalytic triad residues are represented as sticks in magenta.

Figure 9
Docking of Z-LLE-AMC in EspS235A using *AutoDock Vina*. EspS235A is in gold, mature Esp is in cyan and Z-LLE-AMC is in orange. Ser235 of mature Esp and Ala235, Val67 and His250 of EspS235A are represented as sticks and colored magenta.

4.3. *AutoDock*: model of the Z-LLE-AMC–EspS235A complex structure and future direction

Esp cleaves autolysin to inhibit *S. aureus* biofilm formation (Chen *et al.*, 2013), and the substrate Z-LLE-AMC that mimics the autolysin cleavage site was used to test for protease activity. We tried to crystallize the EspS235A–Z-LLE-AMC complex by co-crystallization and by soaking the substrate into EspS235A crystals, but we were not successful.

We used the *AutoDockVina* docking program and proposed a model of the EspS235A–Z-LLE-AMC complex. The docked substrate Z-LLE-AMC was positioned appropriately in the S1 pocket of EspS235A. The N-terminal Val67 amino group was suitably positioned to act as an acceptor of the negative charge of the P1 residue (Glu) side chain of the substrate. The carboxylate group of Glu in Z-LLE-AMC was also at a hydrogen-bond distance from the Ser235/Ala235 residue (Fig. 9). These interactions validate the interaction of Ser235 with the glutamate of Z-LLE-AMC and the subsequent cleavage of the bond between glutamate and AMC, releasing the fluorogenic group AMC. We suspect that the glutamate-harboring loop/region of autolysin would be cleaved similarly, rendering it inactive.

5. Conclusions

S. epidermidis extracellular serine protease (Esp) closely resembles V8 and belongs to a subfamily of chymotrypsin-like serine proteases. It is also a glutamyl endopeptidase that hydrolyzes the peptide bond next to a glutamate residue. Expressed as an inactive zymogen and activated upon losing the pro-peptide segment, Esp has been shown to cleave the autolysin from *S. aureus* and inhibit its biofilm formation. The previously determined mature Esp structure identified an intimate association between the N-terminus (Val67–Gln76) and the substrate-binding S1 pocket (Chen *et al.*, 2013). A structural perspective on the activation of Pro-Esp to Esp, as presented in this study, reveals a mechanism similar to that of the typical serine protease trypsin. The residues on loop 1 and loop 2 in the Pro-Esp structures Esp Δ 55S235A and Esp Δ 55S66V were either flexible or disordered, and the respective oxyanion hole was in a nonfunctional conformation. These crucial segments of the enzyme need to be stabilized before it can function efficiently. The crystal structure of Esp Δ 55S235A also revealed the interaction of the residues in the substrate-binding S1 pocket, Ser235 and His250, with Asn256 and Asp255, respectively, from a symmetry-related molecule, giving us a possible glimpse into the substrate-bound Esp structure.

Furthermore, an N-terminal locked Pro-Esp variant with eight pro-peptide residues did not display any activity, but the crystal structure revealed stabilization of the regions around the S1 pocket (loops 1 and 2 and the oxyanion hole). However, the crystal structure of an N-terminal locked variant with one pro-peptide residue revealed interaction between Ser66 and the active-site serine Ser235, explaining its partial activity despite having a stabilized S1 pocket. Considering the

similarity between the mature structures, we can assume that the activation of the V8 protease zymogen would be similar to that of the Esp zymogen. Overall, we observe that the unusually long pro-peptide segment in bacterial GSEs is flexible and does not associate with the main body of the enzyme, as seen in many cysteine proteases, and the zymogens of GSEs undergo structural changes in the substrate-binding pocket similar to the pro-enzymes of chymotrypsin-like serine proteases upon the loss of their pro-peptide segments. However, the resultant N-termini associate with the substrate-binding pockets of GSEs in a unique way to dictate the specificity and efficiency of the enzyme.

Funding information

This research was funded by National Institute of Allergy and Infectious Diseases grant R01-AI106808 to Narayana Sthanam.

References

- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Bukowski, M., Wladyka, B. & Dubin, G. (2010). *Toxins*, **2**, 1148–1165.
- Calander, A. M., Dubin, G., Potempa, J. & Tarkowski, A. (2008). *FEMS Immunol. Med. Microbiol.* **52**, 267–272.
- Cavarelli, J., Prévost, G., Bourguet, W., Moulinier, L., Chevrier, B., Delagoutte, B., Bilwes, A., Mourey, L., Rifai, S., Piémont, Y. & Moras, D. (1997). *Structure*, **5**, 813–824.
- Chen, C., Krishnan, V., Macon, K., Manne, K., Narayana, S. V. L. & Schneewind, O. (2013). *J. Biol. Chem.* **288**, 29440–29452.
- Chen, Z. & Ruffner, D. E. (1998). *Nucleic Acids Res.* **26**, 1126–1127.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Dancer, S. J., Garratt, R., Saldanha, J., Jhoti, H. & Evans, R. (1990). *FEBS Lett.* **268**, 129–132.
- Demidyuk, I. V., Romanova, D. V., Nosovskaya, E. A., Chestukhina, G. G., Kuranova, I. P. & Kostrov, S. V. (2004). *Protein Eng. Des. Sel.* **17**, 411–416.
- Drapeau, G. R. (1978). *J. Bacteriol.* **136**, 607–613.
- Drapeau, G. R., Boily, Y. & Houmard, J. (1972). *J. Biol. Chem.* **247**, 6720–6726.
- Dubin, G., Chmiel, D., Mak, P., Rakwalska, M., Rzychon, M. & Dubin, A. (2001). *Biol. Chem.* **382**, 1575–1582.
- Edgar, R. C. (2004). *BMC Bioinformatics*, **5**, 113.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Fehlhammer, H., Bode, W. & Huber, R. (1977). *J. Mol. Biol.* **111**, 415–438.
- Fontana, A. (1988). *Biophys. Chem.* **29**, 181–193.
- Fox, T., de Miguel, E., Mort, J. S. & Storer, A. C. (1992). *Biochemistry*, **31**, 12571–12576.
- Gasnov, E. V., Demidyuk, I. V., Shubin, A. V., Kozlovskiy, V. I., Leonova, O. G. & Kostrov, S. V. (2008). *Protein Eng. Des. Sel.* **21**, 653–658.
- Iwase, T., Uehara, Y., Shinji, H., Tajima, A., Seo, H., Takada, K., Agata, T. & Mizunoe, Y. (2010). *Nature (London)*, **465**, 346–349.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kakudo, S., Kikuchi, N., Kitadokoro, K., Fujiwara, T., Nakamura, E., Okamoto, H., Shin, M., Tamaki, M., Teraoka, H., Tsuzuki, H. & Yoshida, N. (1992). *J. Biol. Chem.* **267**, 23782–23788.
- Kawalec, M., Potempa, J., Moon, J. L., Travis, J. & Murray, B. E. (2005). *J. Bacteriol.* **187**, 266–275.
- Khan, A. R. & James, M. N. (1998). *Protein Sci.* **7**, 815–836.
- Klevens, R. M., Morrison, M. A., Nadle, J., Petit, S., Gershman, K., Ray, S., Harrison, L. H., Lynfield, R., Dumyati, G., Townes, J. M., Craig, A. S., Zell, E. R., Fosheim, G. E., McDougal, L. K., Carey, R. B., Fridkin, S. K. & Active Bacterial Core Surveillance (ABCs) MRSA Investigators (2007). *JAMA*, **298**, 1763–1771.
- Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
- Lee, C. Y., Schmidt, J. J., Johnson-Winegar, A. D., Spero, L. & Iandolo, J. J. (1987). *J. Bacteriol.* **169**, 3904–3909.
- Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* **D75**, 861–877.
- Liu, F., Zhao, Z.-S., Ren, Y., Cheng, G., Tang, X.-F. & Tang, B. (2016). *Appl. Microbiol. Biotechnol.* **100**, 10429–10441.
- Massimi, I., Park, E., Rice, K., Müller-Esterl, W., Sauder, D. & McGavin, M. J. (2002). *J. Biol. Chem.* **277**, 41770–41777.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P. & Lopez, R. (2013). *Nucleic Acids Res.* **41**, W597–W600.
- Meijers, R., Blagova, E. V., Levdivkov, V. M., Rudenskaya, G. N., Chestukhina, G. G., Akimkina, T. V., Kostrov, S. V., Lamzin, V. S. & Kuranova, I. P. (2004). *Biochemistry*, **43**, 2784–2791.
- Nemoto, T. K., Ohara-Nemoto, Y., Ono, T., Kobayakawa, T., Shimoyama, Y., Kimura, S. & Takagi, T. (2008). *FEBS J.* **275**, 573–587.
- Neurath, H. & Walsh, K. A. (1976). *Proc. Natl Acad. Sci. USA*, **73**, 3825–3832.
- Nickerson, N. N., Prasad, L., Jacob, L., Delbaere, L. T. & McGavin, M. J. (2007). *J. Biol. Chem.* **282**, 34129–34138.
- Notredame, C., Higgins, D. G. & Heringa, J. (2000). *J. Mol. Biol.* **302**, 205–217.
- Ohara-Nemoto, Y., Ikeda, Y., Kobayashi, M., Sasaki, M., Tajika, S. & Kimura, S. (2002). *Microb. Pathog.* **33**, 33–41.
- Ohara-Nemoto, Y., Ono, T., Shimoyama, Y., Kimura, S. & Nemoto, T. K. (2008). *Biol. Chem.* **389**, 1209–1217.
- Ono, H. K., Omoe, K., Imanishi, K., Iwakabe, Y., Hu, D.-L., Kato, H., Saito, N., Nakane, A., Uchiyama, T. & Shinagawa, K. (2008). *Infect. Immun.* **76**, 4999–5005.
- Ono, T., Ohara-Nemoto, Y., Shimoyama, Y., Okawara, H., Kobayakawa, T., Baba, T. T., Kimura, S. & Nemoto, T. K. (2010). *Biol. Chem.* **391**, 1221–1232.
- O'Toole, P. W. & Foster, T. J. (1987). *J. Bacteriol.* **169**, 3910–3915.
- Park, C. H., Lee, S. J., Lee, S. G., Lee, W. S. & Byun, S. M. (2004). *J. Bacteriol.* **186**, 6457–6464.
- Perona, J. J. & Craik, C. S. (1995). *Protein Sci.* **4**, 337–360.
- Perona, J. J., Hedstrom, L., Rutter, W. J. & Fletterick, R. J. (1995). *Biochemistry*, **34**, 1489–1499.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
- Prasad, L., Leduc, Y., Hayakawa, K. & Delbaere, L. T. J. (2004). *Acta Cryst.* **D60**, 256–259.
- Pustelny, K., Zdzalik, M., Stach, N., Stec-Niemczyk, J., Cichon, P., Czarna, A., Popowicz, G., Mak, P., Drag, M., Salvesen, G. S., Wladyka, B., Potempa, J., Dubin, A. & Dubin, G. (2014). *J. Biol. Chem.* **289**, 15544–15553.
- Qin, X., Singh, K. V., Weinstock, G. M. & Murray, B. E. (2000). *Infect. Immun.* **68**, 2579–2586.
- Rice, K., Peralta, R., Bast, D., de Azavedo, J. & McGavin, M. J. (2001). *Infect. Immun.* **69**, 159–169.
- Rossmann, M. G. (1990). *Acta Cryst.* **A46**, 73–82.

- Rouf, S. M., Ohara-Nemoto, Y., Shimoyama, Y., Kimura, S., Ono, T. & Nemoto, T. K. (2012). *Indian J. Biochem. Biophys.* **49**, 421–427.
- Rühlmann, A., Kukla, D., Schwager, P., Bartels, K. & Huber, R. (1973). *J. Mol. Biol.* **77**, 417–436.
- Stennicke, H. R., Birktoft, J. J. & Breddam, K. (1996). *Protein Sci.* **5**, 2266–2275.
- Suzuki, Y., Yabuta, M. & Ohsuye, K. (1994). *Gene*, **150**, 149–151.
- Svendsen, I. & Breddam, K. (1992). *Eur. J. Biochem.* **204**, 165–171.
- Svendsen, I., Jensen, M. R. & Breddam, K. (1991). *FEBS Lett.* **292**, 165–167.
- Trachuk, L. A., Shcheglov, A. S., Milgotina, E. I. & Chestukhina, G. G. (2005). *Biochimie*, **87**, 529–537.
- Trott, O. & Olson, A. J. (2010). *J. Comput. Chem.* **31**, 455–461.
- Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). *Acta Cryst.* **D60**, 2184–2195.
- Velishaeva, N. S., Gasanov, E. V., Gromova, T. I. & Demidiuk, I. V. (2008). *Bioorg. Khim.* **34**, 786–791.
- Volanakis, J. E. & Narayana, S. V. L. (1996). *Protein Sci.* **5**, 553–564.
- Wiederanders, B., Kaulmann, G. & Schilling, K. (2003). *Curr. Protein Pept. Sci.* **4**, 309–326.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* **D67**, 235–242.