Published in final edited form as:

J Mol Biol. 2011 November 11; 413(5): 985–1000. doi:10.1016/j.jmb.2011.09.028.

Crystal Structure of the Passenger Domain of the *Escherichia coli* Autotransporter EspP

Shekeb Khan^{1,2}, Hira S. Mian^{1,3}, Linda E. Sandercock¹, Nickolay Y. Chirgadze^{1,4}, and Emil F. Pai^{1,2,3,5}

¹Campbell Family Cancer Research Institute, Ontario Cancer Institute, University Health Network, Toronto Medical Discovery Tower, Toronto, Ontario, M5G 1L7, Canada

²Department of Medical Biophysics, University of Toronto, Toronto, Ontario, M5G 1L7, Canada

³Department of Biochemistry, University of Toronto, Toronto, Ontario, M5G 2M9

⁴Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario, M5S 1A8, Canada

⁵Department of Molecular Genetics, University of Toronto, Toronto, Ontario, M5G 1L7, Canada

Abstract

Autotransporters represent a large superfamily of known and putative virulence factors produced by Gram-negative bacteria. They consist of an N-terminal “passenger domain” responsible for the specific effector functions of the molecule and a C-terminal “ β domain” responsible for translocation of the passenger across the bacterial outer membrane. Here we present the 2.5-Å crystal structure of the passenger domain of the extracellular serine protease EspP, produced by the pathogen *Escherichia coli* O157:H7 and a member of the serine protease autotransporters of *Enterobacteriaceae* (SPATEs). Like the previously structurally characterized SPATE passenger domains, the EspP passenger domain contains an extended right-handed parallel β -helix preceded by an N-terminal globular domain housing the catalytic function of the protease. Of note, however, is the absence of a second globular domain protruding from this β -helix. We describe the structure of the EspP passenger domain in the context of previous results and provide an alternative hypothesis for the function of the β -helix within SPATEs.

Introduction

Autotransporters (ATs) form a large superfamily of known and putative virulence proteins secreted by Gram-negative bacteria. They are synthesized as precursors that contain three functional domains: an N-terminal signal sequence, a passenger domain and a C-terminal β -domain. The N-terminal signal sequence is usually about 20-50 amino acid residues in length and targets these proteins to the Sec transport machinery of the cytoplasmic membrane. Translocation across this membrane then proceeds via the Sec pathway utilizing

© 2011 Elsevier Ltd. All rights reserved.

Send correspondence to: Emil F. Pai Ontario Cancer Institute Toronto Medical Discovery Tower, Room 5-358 101 College Street Toronto, Ontario M5G 1L7 Tel: 416-581-7545 pai@hera.med.utoronto.ca .

Accession code Coordinates and structure factors have been deposited in the Research Collaboratory for Structural Bioinformatics PDB under accession code 3SZE.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ATP as an energy source and culminates with the loss of the signal peptide. From within the periplasm, the β -domain, also referred to as the translocator domain, then inserts into the outer membrane as a β -barrel structure and forms a pore through which the rest of the protein (the passenger domain) translocates onto the bacterial cell surface. At this point the passenger domain can remain associated with the outer membrane, via its covalent attachment to the β -domain or in a noncovalent manner following cleavage from the β -domain. Alternatively, as is the case with many ATs, the passenger domain can be released free into the extracellular milieu following cleavage from the β -domain⁸.

The specific effector functions of ATs reside within their secreted passenger domains. The serine protease autotransporters of *Enterobacteriaceae* (SPATEs) represent those ATs produced by *Enterobacteriaceae* where, as the name suggests, the passenger domain functions as a serine protease⁹. Members of this subfamily include, among others, haemoglobin protease (Hbp) and EPEC secreted protein C (EspC)¹² from enteropathogenic *E. coli* (EPEC), the serine protease isoforms of the immunoglobulin A1 proteases (IgAPs) produced by a number of different bacteria, vacuolating autotransporter toxin (Vat)¹⁵ and temperature-sensitive hemagglutinin (Tsh)¹⁶ from avian pathogenic *E. coli* (APEC) as well as EHEC plasmid-encoded autotransporter (EpeA)¹⁷ and extracellular serine protease, plasmid encoded, (EspP) from enterohaemorrhagic *E. coli* (EHEC)¹⁸.

EHEC, especially those of the serotype O157:H7, have been implicated as causative agents in haemorrhagic colitis (HC) and haemolytic uraemic syndrome (HUS). Patients suffering from HC have watery and bloody diarrhoea with mucosal oedema or haemorrhage while HUS is characterized by haemolytic anaemia, consumptive thrombocytopenia and thrombosis in the microvasculature of the kidneys and central nervous system; without treatment, it can lead to renal failure, mental status changes, and a high risk of death. A primary characteristic of EHEC O157:H7 that has been linked to the organism's virulence is the production of the Shiga toxins. However, severe clinical outcomes have also been associated with *E. coli* strains that do not possess such toxins indicating that additional virulence factors might contribute to the pathogenesis of HUS and HC.

In their initial description of EspP, Brunder *et al.*¹⁸ showed that this protease is capable of cleaving porcine pepsin A and, more interestingly, human coagulation factor V, a critical component of the coagulation cascade in human blood. This led the authors to hypothesize that EspP may be an additional virulence factor of EHEC and contribute to the mucosal haemorrhage observed in patients with *E. coli* O157:H7 infection. Almost at the same time, Djafari *et al.*, describing this protease as PssA (protease secreted by Shiga toxin-producing *E. coli*), showed it to be cytotoxic to Vero cells²⁷. In the intervening time, more functional roles have been described for EspP: it cleaves human apolipoprotein A-I²⁸, impairs complement activation by cleaving human complement factors C3/C3b and C5²⁹, cleaves EHEC haemolysin³⁰, forms rope-like fibres with cytopathic and adhesive properties³¹, and mediates intestinal colonization of EHEC O157:H7 in a calf model as well as adherence to primary bovine intestinal epithelial cells³².

The crystallographic structure of the EspP β -domain has previously been described³³. Like other members of the autotransporter family, it forms a 12-stranded β -barrel structure. The passenger domain- β -domain cleavage junction is located inside the barrel pore and experimental data suggest that cleavage is mediated by two amino acid residues lining the pore. As well, the crystallographic structures of the native passenger domains of the SPATEs Hbp¹⁰ from EPEC and IgAP¹³ from *Haemophilus influenzae* have been described. Like the known or predicted structures of passenger domains of most ATs, they contain an extended right-handed β -helical structure that forms a spine to which various other structural features are appended: in Hbp and IgAP, this includes a globular subdomain reminiscent of

the chymotrypsin family of serine proteases, which carries the proteolytic function of the passenger, and a second subdomain (termed domain 2) that adopts a chitinase b-like fold but for which a functional role has not yet been identified³⁷. Also not well defined is the functional role of the β -helical spine of ATs in general. The β -helical structure in ATs, however, through its sequential folding on the bacterial exterior, is hypothesized to provide a mechanism for translocation of the passenger across the bacterial outer membrane.

In order to better understand the functional roles of the EspP passenger, we have determined the crystallographic structure of its Ser263Ala mutant form. We describe the structure in the context of the previous results and provide an alternative hypothesis for the functional role of the β -helix within SPATEs.

Results

Production and characterization of the EspP passenger domain

To study the structural and functional characteristics of the EspP passenger domain, we purified WT EspP passenger (968 residues, numbered 56-1023) from *E. coli* BL21 (DE3) cells transformed with plasmid pB9-5 containing the unmodified native gene *espP*³⁹. SDS-PAGE, analytical size-exclusion chromatography and mass spectrometry revealed the purified passenger domain to be monomeric (data not shown). Functional activity of this domain was confirmed by cleavage of porcine pepsin A and of the tetrapeptide N-suc-Ala-Ala-Pro-Leu-pNA, in agreement with earlier reports. The protein, however, was unstable under normal storage conditions. WT EspP degraded from a native molecular weight of about 105 kDa to two lower weight products, one around 84 kDa and a second around 21 kDa, during the course of a few hours at 25°C as determined by SDS-PAGE. This degradation was inhibited by addition to 2 mM of the serine protease inhibitor phenylmethanesulfonyl fluoride (PMSF). Crystallization attempts of WT EspP with and without PMSF were, however, unsuccessful as well-diffracting crystals were not obtained. We therefore generated a point mutant form of EspP, termed S263A, where the serine residue at position 263 of the molecule has been replaced by an alanine. The S263A passenger domain has abrogated serine protease activity due to removal of the critical hydroxyl moiety from Ser263, thought to be the catalytic serine responsible for proteolytic activity of EspP³⁹. S263A was purified in a manner identical to that for WT EspP, determined to be monomeric and to lack serine protease activity; it also remained stable in solution over the course of many months as determined by SDS-PAGE. Crystallization trials employing S263A yielded protein crystals of sufficient quality to determine the structure of this passenger domain.

Overall structure

The structure of the full-length EspP passenger domain was determined by selenium single-wavelength anomalous dispersion phasing and refined to 2.50 Å resolution with good statistics (Table 1). Each asymmetric unit of the crystal contains one molecule of the passenger domain (~105 kDa). The purified EspP passenger domain contains 968 residues (numbered 56-1023) and can be divided into 2 major structural subdomains: a globular subdomain combining residues 56-313 and 569-630 and a β -helical stalk subdomain spanning the majority of the remaining 648 residues (Figure 1). Residues 581-593 and 1000-1011 are disordered (Fig. 1c) and therefore invisible in the electron density map.

The overall fold of the EspP passenger domain displays high similarity to the folds of the passenger domains of *Escherichia coli* hemoglobin protease (Hbp; PDB ID 1WXR;¹⁰) and *Haemophilus influenzae* immunoglobulin A1 protease (IgAP; PDB ID 3H09;¹³), the two other members of the SPATE family of autotransporters for which high resolution structures

of their passenger domain have been determined. Like Hbp and IgAP, the EspP passenger domain contains an extended β -helical stalk subdomain forming a spine from which protrudes a globular subdomain (termed domain 1 in Hbp) with a fold reminiscent of that of the chymotrypsin family of serine proteases (Fig. 1). Unlike Hbp and IgAP however, the EspP passenger does not contain a second chitinase b-like subdomain (domain 2 in Hbp and IgAP) protruding from the β -helical spine. Instead, a single residue in EspP, namely Ala517, bridges the region occupied by the 77-residue subdomain 2 in Hbp and the 89-residue subdomain 2 in IgAP. If subdomain 2 is indeed involved in substrate recognition by these autotransporters as suggested by Johnson *et al.*¹³, our results show that, unlike Hbp and IgAP, EspP does not utilize such distal structures within the protein fold for substrate recognition; like in other members of the chymotrypsin family, substrate recognition by EspP is likely mediated only by local structural features within and surrounding the active site cleft of the EspP globular subdomain.

The globular subdomain

The N-terminal 258 amino acid residues (residues 56-313; subdomain 1) of the EspP passenger, together with an additional 62 residues (residues 569-630; subdomain 3) further downstream in the primary structure of the passenger, coalesce together to form the globular subdomain of the EspP passenger. However, as with Hbp and IgAP, these two regions of primary structure seem to fold independently and interact over one mostly planar surface with the smaller 62-residue subdomain 3, which forms a helix-turn-helix motif that rests on one face of the structure formed by the larger 258-residue subdomain 1 (Fig. 2). Analysis of this interface using the PISA web tool⁴⁰ reveals a buried surface area of 945 Å² and corresponding solvation free energy of -6.4 kcal/mol, not including the effects of satisfied hydrogen bonds and salt bridges across the interface. This interface is further stabilized by 13 hydrogen bonding interactions and 6 salt bridges. Of note is the presence of a tetramethionine cluster, formed by Met58, Met69 and Met229 from subdomain 1 and Met600 from subdomain 3, which lies at the interface between these two subdomains. According to the PISA web tool, at -3.4 kcal/mol, this cluster accounts for over half of the solvation free energy gain of the interface formation. In contrast, this interface covers an area of 610 Å² in Hbp and 602 Å² in IgAP, lacks any such sidechain clustering, and has a solvation free energy of -4.2 kcal/mol and -2.2 kcal/mol, respectively, indicative of a much weaker interaction between subdomains 1 and 3 in Hbp and IgAP.

Subdomain 3—At 68 amino acid residues in size, subdomain 3 of the EspP passenger is much bulkier than the equivalent region in Hbp (37 residues) and IgAP (34 residues; Fig. 3). However, like in Hbp and IgAP, this region remains devoid of β -strand secondary structure elements, and contains only 20 residues that form α -helical elements in comparison to 10 residues in Hbp and none in IgAP. Consequently, a search of the PDB using the DALI server⁴¹ does not produce any matches between subdomain 3 of these autotransporter passengers and other structures within the PDB.

A further point of difference between subdomain 3 of the EspP passenger and that of Hbp and IgAP is the presence of a disordered loop within subdomain 3 of the EspP passenger, comprising residues Ser581-Gly593. This sequence contains one methionine residue (Met588), from the total of 13 methionines found in the EspP passenger domain, and, interestingly, the two cysteine residues (Cys582 and Cys592), the only cysteines in the entire EspP passenger domain. The oxidation state of these amino acids has been the subject of a number of studies aimed at elucidating the autotransport mechanism – specifically, at determining whether the passenger is translocated across the bacterial outer membrane in an unfolded or in a partially folded state⁴². The mobility of the loop obviously makes it impossible to assign the oxidation state of Cys582 and Cys592. The closest residues

represented by interpretable electron density are Phe580 and Ser594; the relatively short distance of ~ 8.2 Å between their respective C α -atoms certainly allows for Cys582 and Cys592 to be placed at positions (distance of ~ 6 Å between C α -atoms) typical for disulfide bond formation⁴³.

Subdomain 1—Subdomain 1 of the EspP passenger, composed of the N-terminal 258 residues of the protein, shares 27% sequence identity and 39% sequence similarity with the equivalent region (256 residues) in Hbp and 27% sequence identity and 38% sequence similarity with the equivalent region (312 residues) in IgAP. This same region in IgAP shares 31% sequence identity and 43% sequence similarity with the equivalent region in Hbp, and adopts a fold very similar to that in Hbp¹³. Not surprisingly then, subdomain 1 of the EspP passenger, too, adopts a fold very similar to those of Hbp and IgAP.

Subdomain 1 in all of the three structurally characterized SPATE passenger domains adopts a fold reminiscent of the chymotrypsin family of serine proteases. A search of the PDB for structural homologs using the DALI server revealed, other than Hbp and IgAP, that subdomain 1 of the EspP passenger is most similar to bovine pancreatic δ -chymotrypsin (PDBID: 1DLK), yielding a Z-score of 17.3 (Fig. 4). The 178 residues that matched show 12% sequence identity to subdomain 1 of the EspP passenger and could be fitted with a root mean square deviation (RMSD) of 2.5 Å over the main chain. In contrast, subdomain 1 of the EspP passenger yields a Z-score of 25.2 with Hbp, matching on 234 residues with 30% sequence identity and an RMSD of 1.9 Å, and a score of 26.8 with IgAP, matching on 238 residues with 28% sequence identity and an RMSD of 2.1, when analyzed by the DALI server.

Superposition of the structures of δ -chymotrypsin, Hbp, and IgAP with that of EspP produced an excellent match between the catalytic triad of the former three structures with His127, Ser263Ala and Asp156 of EspP, in agreement with earlier predictions that these three residues constitute the catalytic triad of EspP. Closer examination of the region surrounding these residues reveals that EspP is more similar to Hbp than to either IgAP or δ -chymotrypsin. As discussed in the context of the structure of IgAP, the active site cleft of IgAP, which lies at the sideopposite to the β -helix, is more occluded compared to that in Hbp and other members of the chymotrypsin family. In contrast, the active site cleft in EspP is somewhat more exposed than that of Hbp and much more exposed than that of IgAP, suggesting that like Hbp, and unlike IgAP, EspP may be more suited to interacting with a larger surface on its substrate molecules or possibly larger substrates altogether.

Extensive attempts to co-crystallize EspP with bound substrate or inhibitor were unsuccessful. To gain a better insight into substrate specificity of EspP from its crystal structure, the structure of the active site of EspP was compared to that of bovine chymotrypsin in complex with the Lys15Leu mutant variant of bovine pancreatic trypsin inhibitor (BPTI; PDB ID 1P2N; Fig. 5). Using the nomenclature originally proposed by Schechter and Berger⁴⁵ and the scientific knowledge accumulated since then, we know that serine proteases bind to a linear sequence of amino acids on the substrate designated $\text{NH}_3^+ \text{---} \text{P}_n \text{... P}_2, \text{P}_1, \text{P}_1', \text{P}_2' \text{... P}_n' \text{---COO}$ and hydrolyze the peptide bond located between residues P_1 and P_1' . The enzymes in turn have a series of binding pocket subsites designated $\text{S}_n \text{... S}_2, \text{S}_1, \text{S}_1', \text{S}_2' \text{... S}_n'$ each of which accommodates one amino acid residue of the polypeptide substrate, with the P_1^{th} substrate residue being accommodated by the S_1^{th} enzyme subsite. Work by Hedstrom *et al.* and others has also demonstrated that while binding and substrate recognition is a cooperative event involving all enzyme subsites together with the global protein scaffold, subsite S_1 serves as a primary determinant and in at least some cases provides over half of the total association energy for binding and cleavage. With this in mind, and appreciating the limitations of interpreting homology

models when we lack direct experimental evidence in the form of an EspP-substrate or EspP-inhibitor complex crystal structure, we limit our analysis of the EspP binding subsites to subsite S₁.

Comparison of the S₁ subsite of EspP to that of bovine chymotrypsin reveals the former to be both narrower and shallower (Fig. 5). This is in large part due to the presence of Ile286 in EspP filling up the otherwise deeper channel found in bovine chymotrypsin, which has Gly216 at this equivalent position. Furthermore, the S₁ subsite of EspP is slightly more hydrophobic than that of chymotrypsin; Ser189 present at the bottom of the S₁ subsite in chymotrypsin is substituted with Ile286 in the EspP S₁ subsite, Ala258 in EspP replaces Ser190 in chymotrypsin, and the slightly more protruding Ala296 in EspP replaces Gly226 in chymotrypsin. The position of the oxyanion hole, formed by contributions from the backbone amine atoms of Cys191, Gly193, Asp194 and Ser195 in bovine chymotrypsin, is well conserved in EspP through the positioning of Ser259, Gly261, Asp262 and Ser263Ala backbone nitrogen atoms. Taken together, these results suggest that EspP is more suited to cleaving the peptide bond following amino acid residues with smaller and more hydrophobic sidechains than those preferred by chymotrypsin – the S₁ subsite of chymotrypsin is optimized for the large hydrophobic residues tyrosine, tryptophan, phenylalanine, leucine and methionine at the substrate P₁ position. Indeed, the literature suggests that EspP preferentially cleaves the peptide bond immediately N-terminal to leucine, and less frequently threonine²⁹, residues. On the other hand, the S₁ subsite of EspP is slightly wider and much deeper than that found in Hbp (specificity unknown) and IgAP (optimized for proline residues at the substrate P₁ position), clearly indicative of differing substrate P₁ residue preferences between these SPATE autotransporters.

The β -helical stalk subdomain

As with Hbp and IgAP, the most striking structural feature of the EspP passenger is the presence of an extended right-handed β -helix, acting as a spine to which subdomains 1, 3, and 4 as well as various other smaller surface loops are adjoined. In EspP, the prominence of this β -helix is further underscored due to this passenger's lack of a subdomain 2, so that the only major protrusions are subdomains 1, 3 and 4 forming a single globular domain at the N-terminal end of the β -helix.

The EspP β -helix is comprised of 60 β -strands, ranging from as few as 2 residues to 12 residues in length. They are joined together by numerous loops into a right-handed spiral structure with 23 turns and about 23 amino acid residues per turn (Fig. 6a). A large majority of these turns contain three β -strands that themselves hydrogen bond with β -strands in neighbouring turns to form three imperfect parallel β -sheets. The resultant β -helix is reminiscent of a slightly spiralling and slightly kinked imperfect triangular rod (Fig. 6a) and contains a core that is largely but not exclusively filled with aliphatic side chains. Both the N- and C-termini of this helix are capped by a single β -hairpin each that serve to protect the hydrophobic core of the helix from solvent (Fig. 6 b and c). The β -helix of the EspP passenger is one turn shorter than the one found in the Hbp passenger. Sequence and structural alignment indicates that the extra turn in the Hbp β -helix encompasses Hbp residues Lys713-Asp732 and assumes the position equivalent to the junction between rungs 13 and 14, or residues Glu701 and Thr702, of the EspP β -helix. The N-terminal 13 rungs of the Hbp and EspP β -helices, composed of 310 amino acid residues between Asp257-Gly682 in Hbp and 292 amino acid residues between Ser314-Gly668 in EspP, despite having a sequence identity of only 36%, are structurally very similar. They can be superposed on 292 amino acid residues with a DALI Z-score of 25.4 and a root mean square deviation of only 1.6Å. Similarly, the C-terminal 10 rungs of the β -helix in these two autotransporter passengers, spanning 316 residues between Ile733-Asn1048 in Hbp and 322 residues between Thr702-Asn1023 in EspP, have a sequence identity of only 34% but superpose with

a DALI Z-score of 33.2 and a root mean square deviation of again 1.6Å on 310 residues. When compared with the IgAP passenger, the β -helix of the EspP passenger is 4+1/3 turns longer. The IgAP passenger β -helix is best described as lacking the equivalent of rungs 16-191/3, corresponding to amino acids Gly741-Asn860, of the EspP passenger β -helix. The N-terminal 15 rungs of the EspP and IgAP β -helices, spanning 331 residues between Ser314-Asn740 in EspP and 362 residues between Thr338-Glu856 in IgAP, despite having a sequence identity of only 30%, are structurally very similar and superpose on 320 amino acid residues with a DALI Z- score of 35.6 and a root mean square deviation of only 1.7Å. Similarly, the C-terminal 3+2/3 rungs of the β -helix in these two autotransporter passengers, spanning 163 residues between Asn861-Asn1023 in EspP and 158 residues between Asn857-Pro1014 in IgAP have a sequence identity of only 22% but still superpose with a DALI Z-score of 11.5 and a root mean square deviation of only 1.9Å on 121 residues.

A further conserved feature of all SPATE autotransporters, as well as many others, is the presence of a region spanning the C-terminal ~100 amino acid residues of the passenger. Termed the linker or the autochaperone domain (AC), this region is believed to initiate folding of the passenger after translocation across the bacterial outer membrane. The autochaperone region of EspP, with a sequence identity of 43% over 87 amino acid residues when compared to the AC region of Hbp but only 12% over 59 amino acid residues to the AC region of IgAP, adopts a rather similar fold and forms the C-terminal cap on the β -helix of the passenger (Fig. 6 a and b). As with Hbp and IgAP, the AC region of EspP is rich in hydrophobic residues. Gly928, Ala954, Val935, Leu951, Leu992 and Tyr995, numbered according to their positions in EspP, are the only structurally conserved amino acid residues within the folded AC region of these three autotransporters. The latter four of these residues all cluster within one side of the AC, and all six residues with the exception of Tyr995 point into the hydrophobic core of the AC, where they could form the 'folding nucleus' to start the folding process for the whole passenger domain. Indeed, the importance of hydrophobic residues within the AC and the necessity of proper folding of the AC for passenger domain translocation across the outer membrane has been previously described.

Within the surface of the EspP passenger, a concave structure, formed by the passenger's globular subdomain on one side and the right-handed β -helix on the other, generates a large groove, which runs at a ~35 degree angle along the β -helix. It measures ~30Å wide and runs diagonally across the β -helical stalk for ~60Å. It contains a cluster of acidic amino acid residues right at the junction between the globular and the β -helical stalk subdomains (Fig. 7).

Discussion

Whereas the function of the globular subdomain of EspP, and indeed its equivalents in other SPATE autotransporters, is known – they serve as serine endopeptidases– the function of the β -helical stalk subdomain within these autotransporters remains unresolved. Based on sequence alignments and amino acid propensity analysis, >97% of all autotransporters are predicted to contain a β -helical stalk subdomain⁵⁷. Furthermore, while autotransporter passengers traverse the inner cellular membrane of their Gram-negative hosts utilizing the Sec machinery and ATP as an energy source, the energy source allowing for their translocation across the outer cellular membrane remains unknown. Taken together, these observations have led to the proposition that the β -helical stalk serves a structural role and may provide a mechanism to translocate the passenger domain across the bacterial outer membrane by stabilizing the passenger in the extracellular space following its translocation. Specifically, while the early stages of outer membrane translocation remain elusive, Brownian motion of the polypeptide chain is suggested to feed the passenger polypeptide chain, C-terminus first, across the outer membrane through a pore formed by the

autotransporter β -domain³⁸. Concurrently, on the extracellular surface of the outer membrane, the passenger autochaperone region is proposed to serve as a nucleation point to initiate formation of the β -helical stalk subdomain. Recent results suggest that the Bam complex, too, may be involved in the insertion of the β -domain into and translocation of the passenger domain C-terminal region across the bacterial outer membrane. The stiff structural properties of the parallel β -helix together with the hydrogen bonding requirements of the β -strands along the β -helix and its ability to form continuously one β -strand at a time are thought to supply the energy to drive secretion of the polypeptide chain across the bacterial outer membrane. Accordingly, the primary function of the β -helical stalk domain would be to provide a mechanism for translocation of the passenger across the outer membrane.

While convincing arguments have been presented that the β -helical stalk subdomain performs such a function, it certainly cannot be the only role it plays in the process. There exist naturally occurring autotransporters lacking a β -helical stalk subdomain but that nevertheless are fully functional in presenting their native globular passenger domain onto the extracellular surface. The monomeric autotransporter EstA from *Pseudomonas aeruginosa* is an excellent example. While the EstA translocator domain is a typical 12-stranded β -barrel that inserts into the bacterial outer membrane, the EstA passenger domain is dramatically different when compared with the structures of other known passengers; the EstA passenger has a globular fold that predominantly contains α -helices and loops and is covalently attached directly to the translocator domain⁵⁹. The EstA autotransporter has no β -helical structure within it, yet is fully functional in translocating the passenger to the bacterial outer membrane. In addition, numerous studies have demonstrated that engineered autotransporters with their passenger domain replaced in its entirety by non-native and non- β -helical globular surrogates, are fully capable of translocating these surrogates to the extracellular space in functional form⁶¹. In another example also of particular interest to our work, the natural passenger domains of five different autotransporters (EstA, EspP, YpjA, IgAP and AIDA-I) were replaced with small single-chain Anticalin molecules (artificial proteins with a mass of about 20 kDa and derived from human lipocalins)⁶². The Anticalins were successfully translocated across the bacterial outer membrane and were fully functional on the cell exterior. This shows that the entire native passenger domain of EspP up to and including Val998 – in effect removing the entire β -helical stalk of the EspP passenger and most of its autochaperone region (His925-Asn1023) – could be replaced by Anticalins without abrogating translocation of the surrogate passenger to the extracellular surface. The last 25 amino acid residues (Ser999-Asn1024) of the native EspP passenger were left in place to serve as a linker between the Anticalins and the EspP translocator domain thereby ensuring steric accessibility of the surrogate passenger onto the cell surface⁶². These results clearly demonstrate that a β -helical stalk structure is not an absolute requirement for translocation of autotransporter passengers across the outer membrane.

On the other hand, while no such function has yet been described for SPATE autotransporters, many other protein molecules with a parallel β -helical fold are known to be involved in both the recognition and, in some cases, also cleavage of various substrate molecules. One example is found in *Salmonella* sp. bacteriophage P22 tailspike proteins⁶⁵. Surface features within parallel β -helices of these homotrimeric proteins have been identified by x-ray crystallographic analysis and mutational data to directly bind to and cleave O-antigenic repeating units of the bacterial outer membrane lipopolysaccharides. A second example is found in the antifreeze proteins produced by many organisms that inhabit ice-laden environments. These proteins adsorb to the surface of ice crystals and prevent their growth. Structural data over the last 20 years have determined that a significant subset of these proteins adopt a parallel β -helical fold and that surface features within these parallel β -helices mediate adsorption to ice. A third example is seen in the class of enzymes known as

pectinases, including polygalacturonases (PGases) and rhamnogalacturonases (RGases) among others, which are implicated in the degradation of pectic plant cell wall polysaccharides. Many of these enzymes adopt a parallel β -helical fold with a groove almost perpendicular to the helical axis mediating polysaccharide binding and cleavage.

In parallel to these findings, we suggest that the β -helical subdomain of EspP, and indeed those of other autotransporters in general, while involved in the efficient translocation of passengers across the bacterial outer membrane, serves an additional, more specialized role as a mediator of binding to select polymeric structures of the host-organism. When looking at the three-dimensional structure of the EspP passenger, it is tempting to interpret the features of the large groove formed at the junction between the globular and the β -helical stalk subdomains as a surface very well suited to serve as a contact site for an as-yet unidentified molecular binding partner.

In conclusion, we have determined the crystallographic structure of the protease passenger of EspP from the enterohemorrhagic pathogen *Escherichia coli* O157:H7, the first naturally occurring SPATE autotransporter whose fold does not contain a chitinase b-like subdomain (labelled subdomain 2 in Hbp and IgAP). The serine protease function of this protein resides within the globular subdomain of EspP and the catalytic site is at a distance and pointing away from the parallel β -helical stalk domain of this passenger. The precise function of this parallel β -helical stalk subdomain remains unclear but we propose that one of its roles may be in binding to select cellular structures of the host organism. A large cleft formed at the junction between the globular and the β -helical subdomains seems like an excellent candidate for such an interaction, which in turn may play a role in EHEC pathogenesis.

Materials and methods

Site-directed mutagenesis

Plasmid pB9-5 harbouring the gene *espP* from EHEC O157:H7 strain EDL933 has been described previously³⁹. A single S263A mutation was introduced into this gene using the primer 5'-CAT TGC CTC TCA GGG TGA CGC CGG ATC CGC ACT GTT CGT CTA TGA TAA C-3' and its reverse complement and using the Stratagene QuickChange mutagenesis kit (Santa Clara, CA) to generate the plasmid pB9-5-S263A. This mutation abrogates proteolytic activity of the protein by removing the catalytically essential serine residue.

Expression and purification of EspP from culture supernatants

E. coli BL21 (DE3) cells were transformed with recombinant plasmid pB9-5 harbouring the wild-type *espP* gene or pB9-5-S263A harbouring the gene *espP*^{S263A}. A single transformed colony was inoculated in 5 mL LB medium (supplemented with 50 mg/L kanamycin) and grown for 16 hours at 37°C. 4 L LB was inoculated with 4 mL of this primary culture and grown for 12 hours at 37°C. Selenomethionine-labelled protein was expressed similarly but in *E. coli* B834(DE3) and in 1 L of M9 minimal media supplemented with 40 mg/L selenomethionine. The culture was centrifuged at $10,000 \times g$ rcf for 20 minutes at 4°C before the supernatant was collected. It was then supplemented with HEPES, pH 7.4, to 20 mM and NaCl to 1.6 M, filtered through a 0.22 μ m membrane, applied onto a butyl-Sepharose hydrophobic interaction chromatography column (GE Healthcare, Pittsburgh, PA), and eluted with a decreasing linear gradient of NaCl in 20 mM HEPES, pH 7.4. Fractions corresponding to the expected molecular weight were pooled after analyzing for purity on SDS-PAGE, concentrated to 4 mL, and applied onto a Superdex-200 gel filtration column (GE Healthcare, Pittsburgh, PA) pre-equilibrated with 20 mM HEPES, pH 7.4, 300 mM NaCl. Fractions corresponding to the expected molecular weight were again pooled.

after analyzing for purity on SDS-PAGE, concentrated to 10 mg/mL and stored at 4°C, or supplemented with 25% (v/v) glycerol and stored at -70°C. This procedure typically yields ~15 mg of EspP per litre of culture.

Crystallization and data collection

The S263A mutant of EspP (10 mg/mL in 20 mM HEPES pH 7.4, 300 mM NaCl) was crystallized using 18–22% polyethylene glycol 4000, 0.1 M HEPES pH 7.5 and 0.05–0.15 M tris(hydroxymethyl)aminomethane. Crystals were grown at 298 K using the hanging-drop vapour diffusion method, with a starting volume of 4 µL in the drops, composed of 2 µL of protein solution and 2 µL of reservoir solution, and 1 mL of reservoir solution in the well. Large tetragonal crystals belonging to space group $P4_32_12$ with unit cell dimensions $a = b = 88.3$ Å, $c = 311.5$ Å, grew within 1–2 weeks. Crystals were harvested from the drop with a cryoloop, cryoprotected by passing through a 4:1 mixture of Paratone-N and paraffin oil, and flash frozen in liquid nitrogen. Both native and selenomethionine-substituted S263A mutant crystals were prepared in this manner. X-ray diffraction data were collected at the Advanced Photon Source (APS). A 2.5-Å dataset was collected from native crystals on beamline 14-BM-C and a 2.9-Å dataset from selenomethionine-derivative crystals on beamline 17-ID-B. All data were integrated and scaled with HKL2000⁷⁷.

Structure determination and refinement

The structure was determined by the SAD (single-wavelength anomalous dispersion) method⁷⁸ using selenomethionine-containing protein. Initial phase determination and refinement were performed using the Se-SAD data and the programs SHARP and autoSHARP⁷⁹. 12 Se sites were located out of 13 expected. The 3.8-Å resolution map obtained from autoSHARP was of sufficient quality to permit an almost complete unambiguous tracing of the polypeptide backbone, using the program COOT⁸⁰, without further refinement of the heavy atom sites. This model was subsequently used to determine a molecular replacement solution for the 2.5-Å native dataset using the program PHASER⁸¹. The model was further built and refined using an iterative process involving COOT and REFMAC5 in the CCP4 suite using the 2.50-Å resolution native dataset. A final round of TLS (translation/libration/screw)⁸³ refinement was performed in REFMAC5 in the CCP4 suite using 3 groups determined manually per protein molecule: Group 1 consisted of residues 56–313 and 569–629, corresponding to the globular subdomain of the EspP passenger; Group 2 consisted of residues 314–568 and 630–670, corresponding roughly to the N-terminal half of the β-helical stalk subdomain; Group 3 consisted of residues 630–670 and 698–1023, corresponding roughly to the C-terminal half of the β-helical stalk subdomain. Solvent molecules were placed at positions where spherical electron density peaks were found above 1.5 σ in the $2F_o - F_c$ map and where stereochemically reasonable hydrogen bonds were allowed. The final model extends from residue 56, the N-terminal residue of the mature passenger domain, to residue 1023, its C-terminal residue. However, residues 580–593 and 1000–1011 were not represented by interpretable electron density and are therefore missing from the final model. This final model has 85.1% of the residues in the most favourable region, 14.0% in additional allowed regions, and 0.2% in disallowed regions of the Ramachandran plot⁸⁶. The final statistics for data collection and refinement are listed in Table 1.

Acknowledgments

We thank Drs. Werner Brunder and Helge Karch, then at the Institut für Hygiene und Mikrobiologie der Universität Würzburg, Germany, for kindly providing us with the plasmid pB9-5 encoding the gene *espP*. We are grateful to Drs. Yura Lobsanov and Lynne Howell (Hospital for Sick Children, Toronto, ON) for use of their in-house X-ray facility, the staff at Sector 14 of the Advanced Photon Source (Chicago, IL) for help with collection of the native data, Dr. Kevin Battaile from Sector 17 of the Advanced Photon Source for his help in obtaining our Se-SAD data,

and Dr. Robert Lam for support in processing our Se-SAD data. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. DE-AC02-06CH11357. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under grant number RR007707. Use of the IMCA-CAT beamline 17-BM at the Advanced Photon Source was supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with the Center for Advanced Radiation Sources at the University of Chicago. This research was supported by funds from the Canada Research Chairs Program, the Canadian Institutes of Health Research, and the Ontario Research and Development Challenge Fund.

References

1. Dautin N, Bernstein HD. Protein secretion in gram-negative bacteria via the autotransporter pathway. *Annu Rev Microbiol.* 2007; 61:89–112. [PubMed: 17506669]
2. Henderson IR, Navarro-Garcia F, Desvaux M, Fernandez RC, Ala'Aldeen D. Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev.* 2004; 68:692–744. [PubMed: 15590781]
3. Jose J, Jähnig F, Meyer TF. Common structural features of IgA1 protease-like outer membrane protein autotransporters. *Mol Microbiol.* 1995; 18:378–380. [PubMed: 8709857]
4. Pohlner J, Halter R, Beyreuther K, Meyer T. Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. *Nature.* 1987; 325:458–462. [PubMed: 3027577]
5. Henderson IR, Navarro-Garcia F, Nataro JP. The great escape: structure and function of the autotransporter proteins. *Trends Microbiol.* 1998; 6:370–378. [PubMed: 9778731]
6. Oomen CJ, van Ulsen P, van Gelder P, Feijen M, Tommassen J, Gros P. Structure of the translocator domain of a bacterial autotransporter. *EMBO J.* 2004; 23:1257–1266. [PubMed: 15014442]
7. Ieva R, Bernstein HD. Interaction of an autotransporter passenger domain with BamA during its translocation across the bacterial outer membrane. *Proc Natl Acad Sci USA.* 2009; 106:19120–19125. [PubMed: 19850876]
8. Benz I, Schmidt MA. Structures and functions of autotransporter proteins in microbial pathogens. *Int J Med Microbiol.* 2011; 301:461–468. [PubMed: 21616712]
9. Dautin N. Serine Protease Autotransporters of Enterobacteriaceae (SPATEs): Biogenesis and Function. *Toxins.* 2010; 2:1179–1206. [PubMed: 22069633]
10. Otto BR, Sijbrandi R, Luirink J, Oudega B, Heddle JG, Mizutani K, Park S-Y, Tame JRH. Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic *Escherichia coli*. *J Biol Chem.* 2005; 280:17339–17345. [PubMed: 15728184]
11. Otto BR, van Dooren SJ, Nuijens JH, Luirink J, Oudega B. Characterization of a hemoglobin protease secreted by the pathogenic *Escherichia coli* strain EB1. *J Exp Med.* 1998; 188:1091–1103. [PubMed: 9743528]
12. Stein M, Kenny B, Stein MA, Finlay BB. Characterization of EspC, a 110-kilodalton protein secreted by enteropathogenic *Escherichia coli* which is homologous to members of the immunoglobulin A protease-like family of secreted proteins. *J Bacteriol.* 1996; 178:6546–6554. [PubMed: 8932311]
13. Johnson TA, Qiu J, Plaut AG, Holyoak T. Active-site gating regulates substrate selectivity in a chymotrypsin-like serine protease: The structure of *Haemophilus influenzae* immunoglobulin A1 protease. *J Mol Biol.* 2009; 389:559–574. [PubMed: 19393662]
14. Mistry D, Stockley RA. IgA1 protease. *Int J Biochem Cell Biol.* 2006; 38:1244–1248. [PubMed: 16293440]
15. Salvadori MR, Yano T, Carvalho HE, Parreira VR, Gyles CL. Vacuolating cytotoxin produced by avian pathogenic *Escherichia coli*. *Avian Dis.* 2001; 45:43–51. [PubMed: 11332498]
16. Provence DL, Curtiss R. Isolation and characterization of a gene involved in hemagglutination by an avian pathogenic *Escherichia coli* strain. *Infect Immun.* 1994; 62:1369–1380. [PubMed: 8132344]
17. Leyton DL, Sloan J, Hill RE, Doughty S, Hartland EL. Transfer region of pO113 from enterohemorrhagic *Escherichia coli*: similarity with R64 and identification of a novel plasmid-encoded autotransporter, EpeA. *Infect Immun.* 2003; 71:6307–6319. [PubMed: 14573650]

18. Brunder W, Schmidt H, Karch H. EspP, a novel extracellular serine protease of enterohaemorrhagic *Escherichia coli* O157:H7 cleaves human coagulation factor V. *Mol Microbiol.* 1997; 24:767–778. [PubMed: 9194704]
19. Riley LW, Remis RS, Helgerson SD, McGee HB, Wells JG, Davis BR, Hebert RJ, Olcott ES, Johnson LM, Hargrett NT, Blake PA, Cohen ML. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N Engl J Med.* 1983; 308:681–685. [PubMed: 6338386]
20. Pennington H. *Escherichia coli* O157. *Lancet.* 2010; 376:1428–1435. [PubMed: 20971366]
21. Griffin, PM. *Escherichia coli* O157:H7 and other enterohemorrhagic *Escherichia coli*. In: Blaser, MJ.; Smith, PD.; Ravdin, JI.; Greenberg, HB.; Guerrant, RL., editors. *Infections of the Gastrointestinal Tract.* Raven Press Ltd.; New York: 1995. p. 739-761.
22. Buchanan RL, Doyle MP. Foodborne disease significance of *Escherichia coli* O157:H7 and other enterohemorrhagic *E. coli*. *Food Technol.* 1997; 51:69–76.
23. Tarr PI, Gordon CA, Chandler WL. Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. *Lancet.* 2005; 365:1073–1086. [PubMed: 15781103]
24. Brockmeyer J, Bielaszewska M, Fruth A, Bonn ML, Mellmann A, Humpf H-U, Karch H. Subtypes of the plasmid-encoded serine protease EspP in Shiga toxin-producing *Escherichia coli*: distribution, secretion, and proteolytic activity. *Appl Environ Microbiol.* 2007; 73:6351–6359. [PubMed: 17704265]
25. Mellmann A, Bielaszewska M, Zimmerhackl LB, Prager R, Harmsen D, Tschäpe H, Karch H. Enterohemorrhagic *Escherichia coli* in human infection: in vivo evolution of a bacterial pathogen. *Clin Infect Dis.* 2005; 41:785–792. [PubMed: 16107974]
26. Schmidt H, Scheef J, Huppertz HI, Frosch M, Karch H. *Escherichia coli* O157:H7 and O157:H(-) strains that do not produce Shiga toxin: phenotypic and genetic characterization of isolates associated with diarrhea and hemolytic-uremic syndrome. *J Clin Microbiol.* 1999; 37:3491–3496. [PubMed: 10523540]
27. Djafari S, Ebel F, Deibel C, Krämer S, Hudel M, Chakraborty T. Characterization of an exported protease from Shiga toxin-producing *Escherichia coli*. *Mol Microbiol.* 1997; 25:771–784. [PubMed: 9379905]
28. Schmidt H, Zhang WL, Hemmrich U, Jelacic S, Brunder W, Tarr PI, Dobrindt U, Hacker J, Karch H. Identification and characterization of a novel genomic island integrated at selC in locus of enterocyte effacement-negative, Shiga toxin-producing *Escherichia coli*. *Infect Immun.* 2001; 69:6863–6873. [PubMed: 11598060]
29. Orth D, Ehrlenbach S, Brockmeyer J, Khan AB, Huber G, Karch H, Sarg B, Lindner H, Würzner R. EspP, a serine protease of enterohemorrhagic *Escherichia coli*, impairs complement activation by cleaving complement factors C3/C3b and C5. *Infect Immun.* 2010; 78:4294–4301. [PubMed: 20643852]
30. Brockmeyer J, Aldick T, Soltwisch J, Zhang W, Tarr PI, Weiss A, Dreisewerd K, Muthing J, Bielaszewska M, Karch H. Enterohaemorrhagic *Escherichia coli* haemolysin is cleaved and inactivated by serine protease EspPa. *Environ Microbiol.* 2011; 13:1327–1341. [PubMed: 21352460]
31. Xicohtencatl-Cortes J, Saldana Z, Deng W, Castaneda E, Freer E, Tarr PI, Finlay BB, Puente JL, Giron JA. Bacterial macroscopic rope-like fibers with cytopathic and adhesive properties. *J Biol Chem.* 2010; 285:32336–32342. [PubMed: 20688909]
32. Dziva F, Mahajan A, Cameron P, Currie C, McKendrick IJ, Wallis TS, Smith DGE, Stevens MP. EspP, a Type V-secreted serine protease of enterohaemorrhagic *Escherichia coli* O157:H7, influences intestinal colonization of calves and adherence to bovine primary intestinal epithelial cells. *FEMS Microbiol Lett.* 2007; 271:258–264. [PubMed: 17451446]
33. Barnard TJ, Dautin N, Lukacik P, Bernstein HD, Buchanan SK. Autotransporter structure reveals intra-barrel cleavage followed by conformational changes. *Nat Struct Mol Biol.* 2007; 14:1214–1220. [PubMed: 17994105]
34. Dautin N, Barnard TJ, Anderson DE, Bernstein HD. Cleavage of a bacterial autotransporter by an evolutionarily convergent autocatalytic mechanism. *EMBO J.* 2007; 26:1942–1952. [PubMed: 17347646]

35. Emsley P, Charles IG, Fairweather NF, Isaacs NW. Structure of *Bordetella pertussis* virulence factor P.69 pertactin. *Nature*. 1996; 381:90–92. [PubMed: 8609998]
36. Gangwer KA, Mushrush DJ, Stauff DL, Spiller B, McClain MS, Cover TL, Lacy DB. Crystal structure of the *Helicobacter pylori* vacuolating toxin p55 domain. *Proc Natl Acad Sci USA*. 2007; 104:16293–16298. [PubMed: 17911250]
37. Nishimura K, Yoon YH, Kurihara A, Unzai S, Lührink J, Park SY, Tame JRH. Role of domains within the autotransporter Hbp/Tsh. *Acta Crystallogr D Biol Crystallogr*. 2010; 66:1295–300. [PubMed: 21123869]
38. Klauser T, Pohlner J, Meyer TF. Selective extracellular release of cholera toxin B subunit by *Escherichia coli*: dissection of *Neisseria* Iga beta-mediated outer membrane transport. *EMBO J*. 1992; 11:2327–2335. [PubMed: 1600950]
39. Brunder W, Schmidt H, Karch H. EspP, a novel extracellular serine protease of enterohaemorrhagic *Escherichia coli* O157:H7 cleaves human coagulation factor V. *Mol Microbiol*. 1997; 24:767–778. [PubMed: 9194704]
40. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007; 372:774–797. [PubMed: 17681537]
41. Holm L, Rosenström P. Dali server: conservation mapping in 3D. *Nucleic Acids Res*. 2010; 38:W545–W549. [PubMed: 20457744]
42. Skillman KM, Barnard TJ, Peterson JH, Ghirlando R, Bernstein HD. Efficient secretion of a folded protein domain by a monomeric bacterial autotransporter. *Mol Microbiol*. 2005; 58:945–958. [PubMed: 16262782]
43. Goldblum A, Rein R. Geometrical analysis of Cys-Cys bridges in proteins and their prediction from incomplete structural information. *Int J Pept Protein Res*. 1987; 30:784–793. [PubMed: 11542515]
44. Brockmeyer J, Spelten S, Kuczius T, Bielaszewska M, Karch H. Structure and function relationship of the autotransport and proteolytic activity of EspP from Shiga toxin-producing *Escherichia coli*. *PLoS ONE*. 2009; 4:e6100. [PubMed: 19568421]
45. Schechter I, Berger A. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun*. 1967; 27:157–162. [PubMed: 6035483]
46. Hedstrom L, Szilagyi L, Rutter WJ. Converting trypsin to chymotrypsin: the role of surface loops. *Science*. 1992; 255:1249–1253. [PubMed: 1546324]
47. Perona JJ, Craik CS. Structural basis of substrate specificity in the serine proteases. *Protein Sci*. 1995; 4:337–360. [PubMed: 7795518]
48. Perona JJ, Hedstrom L, Rutter WJ, Fletterick RJ. Structural origins of substrate discrimination in trypsin and chymotrypsin. *Biochemistry*. 1995; 34:1489–1499. [PubMed: 7849008]
49. Perona JJ, Craik CS. Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. *J Biol Chem*. 1997; 272:29987–29990. [PubMed: 9374470]
50. Helland R, Czapińska H, Leiros I, Olufsen M, Otlewski J, Smalås AO. Structural consequences of accommodation of four non-cognate amino acid residues in the S1 pocket of bovine trypsin and chymotrypsin. *J Mol Biol*. 2003; 333:845–861. [PubMed: 14568540]
51. Ohnishi Y, Nishiyama M, Horinouchi S, Beppu T. Involvement of the COOH-terminal pro-sequence of *Serratia marcescens* serine protease in the folding of the mature enzyme. *J Biol Chem*. 1994; 269:32800–32806. [PubMed: 7806503]
52. Oliver DC, Huang G, Nodel E, Pleasance S, Fernandez RC. A conserved region within the *Bordetella pertussis* autotransporter BrkA is necessary for folding of its passenger domain. *Mol Microbiol*. 2003; 47:1367–1383. [PubMed: 12603741]
53. Kühnel K, Diezmann D. Crystal Structure of the Autochaperone Region from the *Shigella flexneri* Autotransporter IcsA. *J Bacteriol*. 2011; 193:2042–2045. [PubMed: 21335457]
54. Peterson JH, Tian P, Ieva R, Dautin N, Bernstein HD. Secretion of a bacterial virulence factor is driven by the folding of a C-terminal segment. *Proc Natl Acad Sci USA*. 2010; 107:17739–17744. [PubMed: 20876094]
55. Velarde JJ, Nataro JP. Hydrophobic residues of the autotransporter EspP linker domain are important for outer membrane translocation of its passenger. *J Biol Chem*. 2004; 279:31495–31504. [PubMed: 15151995]

56. Renn JP, Clark PL. A conserved stable core structure in the passenger domain beta-helix of autotransporter virulence proteins. *Biopolymers*. 2008; 89:420–427. [PubMed: 18189304]
57. Junker M, Schuster CC, McDonnell AV, Sorg KA, Finn MC, Berger B, Clark PL. Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proc Natl Acad Sci USA*. 2006; 103:4918–4923. [PubMed: 16549796]
58. Thanassi DG, Stathopoulos C, Karkal A, Li H. Protein secretion in the absence of ATP: the autotransporter, two-partner secretion and chaperone/usher pathways of gram-negative bacteria (review). *Mol Membr Biol*. 2005; 22:63–72. [PubMed: 16092525]
59. van den Berg B. Crystal structure of a full-length autotransporter. *J Mol Biol*. 2010; 396:627–633. [PubMed: 20060837]
60. Wilhelm S, Tommassen J, Jaeger KE. A novel lipolytic enzyme located in the outer membrane of *Pseudomonas aeruginosa*. *J Bacteriol*. 1999; 181:6977–6986. [PubMed: 10559163]
61. Jose J, Meyer TF. The autodisplay story, from discovery to biotechnical and biomedical applications. *Microbiol Mol Biol Rev*. 2007; 71:600–619. [PubMed: 18063719]
62. Binder U, Matschiner G, Theobald I, Skerra A. High-throughput sorting of an Anticalin library via EspP-mediated functional display on the *Escherichia coli* cell surface. *J Mol Biol*. 2010; 400:783–802. [PubMed: 20630471]
63. Petersen TN, Kauppinen S, Larsen S. The crystal structure of rhamnogalacturonase A from *Aspergillus aculeatus*: a right-handed parallel beta helix. *Structure*. 1997; 5:533–544. [PubMed: 9115442]
64. Jenkins J, Pickersgill R. The architecture of parallel beta-helices and related folds. *Prog Biophys Mol Biol*. 2001; 77:111–175. [PubMed: 11747907]
65. Steinbacher S, Seckler R, Miller S, Steipe B, Huber R, Reinemer P. Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer. *Science*. 1994; 265:383–386. [PubMed: 8023158]
66. Steinbacher S, Baxa U, Miller S, Weintraub A, Seckler R, Huber R. Crystal structure of phage P22 tailspike protein complexed with *Salmonella* sp. O-antigen receptors. *Proc Natl Acad Sci USA*. 1996; 93:10584–10588. [PubMed: 8855221]
67. Baxa U, Steinbacher S, Miller S, Weintraub A, Huber R, Seckler R. Interactions of phage P22 tails with their cellular receptor, *Salmonella* O-antigen polysaccharide. *Biophys J*. 1996; 71:2040–2048. [PubMed: 8889178]
68. Steinbacher S, Miller S, Baxa U, Budisa N, Weintraub A, Seckler R, Huber R. Phage P22 tailspike protein: crystal structure of the head-binding domain at 2.3 Å, fully refined structure of the endorhamnosidase at 1.56 Å resolution, and the molecular basis of O-antigen recognition and cleavage. *J Mol Biol*. 1997; 267:865–880. [PubMed: 9135118]
69. Raymond JA, DeVries AL. Adsorption inhibition as a mechanism of freezing resistance in polar fishes. *Proc Natl Acad Sci USA*. 1977; 74:2589–2593. [PubMed: 267952]
70. Garnham CP, Campbell RL, Davies PL. From the Cover: Anchored clathrate waters bind antifreeze proteins to ice. *Proc Natl Acad Sci USA*. 2011; 108:7363–7367. [PubMed: 21482800]
71. Graether SP, Kuiper MJ, Gagné SM, Walker VK, Jia Z, Sykes BD, Davies PL. Beta-helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect. *Nature*. 2000; 406:325–328. [PubMed: 10917537]
72. Cho SW, Lee S, Shin W. The X-ray structure of *Aspergillus aculeatus* polygalacturonase and a modeled structure of the polygalacturonase-octagalacturonate complex. *J Mol Biol*. 2001; 311:863–878. [PubMed: 11518536]
73. Abbott DW, Boraston AB. Structural biology of pectin degradation by Enterobacteriaceae. *Microbiol Mol Biol Rev*. 2008; 72:301–316. [PubMed: 18535148]
74. Scavetta RD, Herron SR, Hotchkiss AT, Kita N, Keen NT, Benen JA, Kester HC, Visser J, Jurnak F. Structure of a plant cell wall fragment complexed to pectate lyase C. *Plant Cell*. 1999; 11:1081–1092. [PubMed: 10368179]
75. Pickersgill R, Smith D, Worboys K, Jenkins J. Crystal structure of polygalacturonase from *Erwinia carotovora* ssp. *carotovora*. *J Biol Chem*. 1998; 273:24660–24664. [PubMed: 9733763]
76. Fries M, Ihrig J, Brocklehurst K, Shevchik VE, Pickersgill RW. Molecular basis of the activity of the phytopathogen pectin methylesterase. *EMBO J*. 2007; 26:3879–3887. [PubMed: 17717531]

77. Otwinowski Z, Minor W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Meth Enzymol.* 1997; 276:307–326.
78. Wang BC. Resolution of phase ambiguity in macromolecular crystallography. *Meth Enzymol.* 1985; 115:90–112. [PubMed: 4079800]
79. Vonrhein C, Blanc E, Roversi P, Bricogne G. Automated structure solution with autoSHARP. *Methods Mol Biol.* 2007; 364:215–230. [PubMed: 17172768]
80. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* 2004; 60:2126–32. [PubMed: 15572765]
81. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr.* 2007; 40:658–674. [PubMed: 19461840]
82. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr.* 1997; 53:240–55. [PubMed: 15299926]
83. Winn MD, Murshudov GN, Papiz MZ. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Meth Enzymol.* 2003; 374:300–321. [PubMed: 14696379]
84. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr.* 1994; 50:760–3. [PubMed: 15299374]
85. Potterton E, Briggs P, Turkenburg M, Dodson E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr.* 2003; 59:1131–7. [PubMed: 12832755]
86. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *J Mol Biol.* 1963; 7:95–99. [PubMed: 13990617]

Research Highlights

Crystal structure of the passenger domain of the *E. coli* autotransporter EspP identified as member of SPATE protein family β -helix stalk proposed to engage in binding to extended structures in host cells

Figure 1. Crystal structure of the EspP passenger domain

a) Linear schematic representation of the EspP primary sequence, with the globular subdomain coloured red (subdomain 1 or SD1, residues 56-313) and green (subdomain 3 or SD3, residues 569-630), subdomain 4 (or SD4, residues 671-699) coloured yellow and the β -helical stalk subdomain coloured blue. The signal sequence and the translocator domain (β -domain) are coloured in gray. For ease of comparison with the other structurally known SPATE autotransporter passenger domains, the established nomenclature for subdomains is used although EspP does not contain the equivalent of subdomain 2. b) Wall-eyed stereo ribbon diagram of the EspP passenger coloured as in (a). Encircled in light blue and rendered as stick models are the catalytic triad residues (His127, Asp156 and Ser263Ala) of the passenger. c) Ribbon diagram of the EspP passenger rotated 90° clockwise about a vertical axis relative to (b). Represented as dotted loops are residues 581-593 (green) and 1000-1011 (blue); these residues are disordered in the crystal and not visible in the electron density map. d) Ribbon diagram of the passenger domain of the SPATE autotransporter Hbp (PDB ID: 1WXR) shown in the same orientation and to the same scale as the EspP passenger domain in (c). Subdomain 2 of the Hbp passenger is labeled and coloured purple. e) Wall-eyed stereo representation of the electrostatic surface of the EspP passenger oriented as in (b), with red indicating electronegative potential and blue indicating electropositive potential. Note the patch of negative charge at the bottom of the groove where the globular and β -helical domains meet.

Figure 2. Globular subdomain architecture of the EspP passenger

a) Subdomain 3 (green) forms a helix-turn-helix motif that rests on one face of subdomain 1 (red). Residues 581-593 within subdomain 3 are disordered in the crystal structure and are presented here as a green dotted loop. b) A tetramethionine cluster lies at the interface of subdomains 1 and 3, likely providing further structural stability to the globular subdomain. Interatomic distances between the sulfur atoms of the methionine cluster are indicated by yellow dashed lines and marked in Å units.

Figure 3. Subdomains 3 of (a) EspP with that of (b) Hbp and (c) IgAP for comparison
Subdomains 3 of EspP, Hbp and IgAP are presented as cartoons and coloured green with the remainder of each molecule rendered as a faded gray molecular surface.

Figure 4. Subdomains 1 of (a) EspP with that of (b) Hbp and (c) IgAP and (d) δ -chymotrypsin for comparison

Subdomains 1 of EspP, Hbp and IgAP and the structure of δ -chymotrypsin are presented as cartoons, with α -helices coloured red and β -strands coloured dark blue. The sidechains of the serine, histidine and aspartate residues forming the catalytic triad of each protease are presented as stick models and coloured with carbon atoms in silver, nitrogen atoms in blue and oxygen atoms in red. The bound inhibitor present in the crystal structure of δ -chymotrypsin is presented as a stick model and coloured purple. All images are rendered from the same view. To orient the reader, subdomain 1 of the EspP passenger is presented in the context of subdomains 3 (green cartoon) and 4 (yellow surface) and the β -helical spine subdomain (light blue surface) of EspP.

Figure 5. Comparison of the putative S1 binding pocket subsite of EspP with the S1 subsites of bovine chymotrypsin, Hbp, and IgAP

a) Wall-eyed stereo view of the S1 binding pocket subsite of bovine trypsin in complex with bovine pancreatic trypsin inhibitor (BPTI; PDB ID 1P2N). b) Bovine chymotrypsin was matched onto EspP by least-squares fitting residues 56-58, 101-103, 137-139, 183-196 and 212-229 in bovine chymotrypsin with residues 126-128, 155-157, 202-204, 247-264 and 282-299 in EspP. This match produces a root mean square deviation of 1.5 Å over all backbone atoms. The position of BPTI from its complex with bovine chymotrypsin was then transposed onto the structure of EspP to highlight the putative S1 binding pocket of EspP. This panel is a wall-eyed stereo view of BPTI as bound in bovine trypsin superimposed onto the putative S1 binding pocket of EspP. (a and b) BPTI (green) is depicted from positions P4-P4' (Gly12-Ile19) with the residue at position P1 (Leu15) together with the backbone atoms spanning positions P2-P2' (Cys14-Arg17) rendered as stick models. The catalytic triad of bovine chymotrypsin and EspP are also rendered as stick models and labeled. Protein residues surrounding the inhibitor position P1 (Leu15) are rendered as line models and labeled. c-f) Electrostatic surface representation of the substrate binding pocket of (c) bovine chymotrypsin in complex with BPTI (BPTI rendered as in (a and b)), (d) EspP, (e) Hbp and (f) IgAP, with red indicating electronegative potential and blue indicating electropositive potential. The S1 binding pocket subsites in (c-f) are highlighted by yellow circles.

Figure 6. The β -helical stalk subdomain of EspP

a) The β -helical stalk subdomain of EspP is presented as a ribbon diagram with α -helices coloured yellow and β -strands coloured blue. Subdomains 1, 3 and 4 are presented as faded gray molecular surfaces. Dotted loops represent residues disordered in the crystal structure: In EspP these are residues 581-593 (faded gray) and 1000-1011 (orange). The whole autochaperone (AC) region of EspP (residues 925-1023) is coloured orange. Every fifth turn of the β -helix along with the first (1) and last (23) turns are marked by green dots and numbered at the far left. b) Close-up view of the EspP AC region and comparison to the same region of Hbp (PDB ID: 1WXR) and IgAP (PDB ID: 3H09). One turn of the β -helix immediately preceding the AC region in each autotransporter is shown for reference. Residues 938-940 (light blue dotted loop) and 1025-1030 (orange dotted loop) in Hbp and residues 990-1014 (orange dotted loop) in IgAP are not present in the respective models. Shown in green with side chains rendered as sticks, are those residues that are absolutely conserved between the AC regions of EspP, Hbp and IgAP. c) Close-up view of the EspP β -helical stalk N-terminal cap and comparison to that of Hbp and IgAP. Shown are the first three turns of the β -helix in each protein, with α -helices coloured yellow and β -strands coloured blue. In white are the last 4-5 residues of subdomain 1 of each protein leading into the β -helix.

Figure 7. Proposed binding cleft within the EspP passenger domain

The EspP passenger molecular surface is presented with the same colour scheme as in Figure 1a and as an electrostatic surface. The proposed binding cleft within the EspP passenger is marked with dashed lines.

Table 1

Data collection and refinement statistics.

	Se-SAD ^a	Native ^a
Data collection		
X-ray wavelength (Å)	0.9793581	0.90020
Space group	<i>P</i> 4 ₃ 2 ₁ 2	<i>P</i> 4 ₃ 2 ₁ 2
Unit-cell parameters (Å)		
<i>a</i> and <i>b</i>	88.29	88.34
<i>c</i>	311.29	311.53
Resolution range (Å)	50–2.87 (2.97–2.87)	30–2.50 (2.56–2.50)
No. of measured reflections	407947	303416
No. of unique reflections	29182 (2836)	43730 (2848)
Completeness (%)	99.8 (99.5)	99.9 (99.8)
Redundancy	14.0 (13.0)	6.9 (7.2)
<i>I</i> / $\sigma(I)$	23.0 (12.1)	14.0 (3.9)
<i>R</i> _{merge} ^b	8.8 (22.4)	10.9 (48.6)
Model refinement		
<i>R</i> _{work} / <i>R</i> _{free} ^c	–	0.188 / 0.231
No. of non-H atoms / average <i>B</i> factor (Å ²)		
Protein	–	7252 / 29.5
Waters	–	324 / 28.0
R.m.s. deviations from ideal geometry		
Bond lengths (Å)	–	0.008
Bond angles (°)	–	1.11
Ramachandran plot		
Most favourable (%)	–	85.1
Allowed (%)	–	14.0
Generously allowed (%)	–	0.7
Disallowed (%)	–	0.2

^aValues in parenthesis are for the outer shell.^b $R = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where *I*(*hkl*) is the intensity of reflection *hkl*, $\sum_{hkl}$ is the sum over all reflections and $\sum_i$ is the sum over *i* measurements of reflection *hkl*.^c $R = \sum_{hkl} |F_{obs} - F_{calc}| / \sum_{hkl} F_{obs}$, where *F*_{obs} and *F*_{calc} are the observed and calculated structure-factor amplitudes, respectively. *R*_{free} is calculated for a randomly chosen 5% of reflections that were not used for structure refinement and *R*_{work} is calculated for the remaining reflections.