

The structure of *Serratia marcescens* Lip, a membrane-bound component of the type VI secretion system

Vincenzo A. Rao, Sharon M. Shepherd, Grant English, Sarah J. Coulthurst and William N. Hunter*

College of Life Sciences, University of Dundee,
Dundee DD1 5EH, Scotland

Correspondence e-mail:
w.n.hunter@dundee.ac.uk

Lip is a membrane-bound lipoprotein and a core component of the type VI secretion system found in Gram-negative bacteria. The structure of a Lip construct (residues 29–176) from *Serratia marcescens* (*SmLip*) has been determined at 1.92 Å resolution. Experimental phases were derived using a single-wavelength anomalous dispersion approach on a sample cocrystallized with iodide. The membrane localization of the native protein was confirmed. The structure is that of the globular domain lacking only the lipoprotein signal peptide and the lipidated N-terminus of the mature protein. The protein fold is dominated by an eight-stranded β -sandwich and identifies *SmLip* as a new member of the transthyretin family of proteins. Transthyretin and the only other member of the family fold, 5-hydroxyisourate hydrolase, form homotetramers important for their function. The asymmetric unit of *SmLip* is a tetramer with 222 symmetry, but the assembly is distinct from that previously noted for the transthyretin protein family. However, structural comparisons and bacterial two-hybrid data suggest that the *SmLip* tetramer is not relevant to its role as a core component of the type VI secretion system, but rather reflects a propensity for *SmLip* to participate in protein–protein interactions. A relatively low level of sequence conservation amongst Lip homologues is noted and is restricted to parts of the structure that might be involved in interactions with physiological partners.

Received 21 September 2011

Accepted 2 November 2011

PDB Reference: Lip, 4a1r.

1. Introduction

Protein secretion systems are critical to the virulence and host-interaction processes of Gram-negative pathogens. Different bacterial species possess different combinations of one or more specialized proteinaceous machines that secrete toxins, adhesins, hydrolytic enzymes and proteins able to manipulate eukaryotic signalling pathways (Gerlach & Hensel, 2007; Holland, 2010). The most recently discovered system, the type VI secretion system (T6SS), is present in many Gram-negative bacteria and is implicated in virulence in important human pathogens including *Pseudomonas aeruginosa* (Cascales, 2008; Filloux *et al.*, 2008; Jani & Cotter, 2010). It has also been shown to contribute to the virulence of economically significant animal and plant pathogens (Liu *et al.*, 2008; Blondel *et al.*, 2010; Sarris *et al.*, 2010). Some T6SSs appear to target other bacterial cells instead of, or in addition to, eukaryotic cells (Hood *et al.*, 2010; MacIntyre *et al.*, 2010; Murdoch *et al.*, 2011). This suggests that T6SSs may contribute to allowing pathogens to proliferate in polymicrobial infection sites and/or to persist in different environmental reservoirs (Schwarz *et al.*, 2010; Murdoch *et al.*, 2011). *Serratia marcescens* is an opportunistic

pathogen, a significant cause of hospital-acquired infections and an important reservoir of antibiotic-resistance determinants in the clinical environment (Hejazi & Falkner, 1997). It is also a tractable model organism in which to dissect the structure–function relationships in the T6SS (Murdoch *et al.*, 2011).

Studies of the T6SS have started to reveal information on the components and the biological role of this recently discovered system (Cascales, 2008; Filloux *et al.*, 2008; Pukatzki *et al.*, 2009; Bönenmann *et al.*, 2010). T6SSs are large multiprotein complexes encoded on variable gene clusters characterized by the presence of genes encoding 13 ‘core’ components. These are thought to form the basic secretion apparatus, which is coupled with ‘accessory’ components that are conserved across many or only a few systems. Key core components include the putative extracellular Hcp/VgrG assembly, which is thought to form a cell-puncturing device similar to that of bacteriophage tail structures (Pukatzki *et al.*, 2009). There are a number of predicted cytoplasmic proteins (*e.g.* an ATPase called ClpV) and several inner membrane proteins (*e.g.* IcmF and IcmH). Additionally, and the subject of this work, the only outer membrane component reported to date is a periplasmic-facing outer membrane lipoprotein (Lip; Aschtgen *et al.*, 2008).

Genetic studies indicate that in *S. marcescens* this lipoprotein (*SmLip*) makes an essential contribution to the basic function of the T6SS and to T6SS-dependent antibacterial killing activity (Murdoch *et al.*, 2011). We now report the high-resolution structure of *SmLip* determined following phase determination using single-wavelength anomalous dispersion (SAD) measurements based on the scattering properties of iodide ions. The localization of the protein in *S. marcescens* itself and bacterial two-hybrid data are reported to investigate the propensity for self-association. The structure reveals a remarkable similarity to transthyretin, a vertebrate hormone-distribution protein, and comparisons suggest which parts of *SmLip* may be involved in protein–protein interactions with partner components of the T6SS.

2. Methods

2.1. Protein expression and purification

The *S. marcescens lip* gene (SMA2252; Murdoch *et al.*, 2011) encoding amino-acid residues 30–176 was amplified from genomic DNA (strain Db10) using the forward primer 5′-catatgCCAAAAGCGTGCCGTCGCGTTACAG-3′ and the reverse primer 5′-ggatccTCAGTCGACCTTTTTTACGGGGCGCAGGC-3′ (the lower-case sequences correspond to the *NdeI/BamHI* restriction sites used for cloning). The PCR product was ligated into PCR-BluntII-TOPO using the Zero Blunt TOPO Cloning Kit (Invitrogen) and then cloned into a pET15b (Novagen) cloning vector modified to encode a tobacco etch virus (TEV) protease cleavage site in place of the thrombin protease cleavage site. The construct was verified by DNA sequencing (DNA Sequencing Unit, University of Dundee).

The recombinant protein was produced in *Escherichia coli* BL21 (DE3) pLysS cells (Stratagene). Cultures were grown for 3 h at 310 K in auto-induction medium (Studier, 2005) supplemented with 50 µg l⁻¹ carbenicillin before overnight growth at 295 K. Cells were harvested by centrifugation (3500g at 277 K for 30 min). The cell pellet was resuspended in buffer *A* (25 mM Tris–HCl pH 7.5, 500 mM NaCl, 20 mM imidazole pH 8.5) supplemented with an EDTA-free Protease Inhibitor Cocktail Tablet (Roche) and 0.2 mg DNase I (Sigma–Aldrich). Cells were lysed using a continuous-flow cell disrupter (Constant Systems) at 207 MPa and cell debris was removed following centrifugation (40 000g at 277 K for 30 min). *SmLip* was purified using nickel-affinity chromatography with a 5 ml HisTrap HP column (GE Healthcare) pre-charged with Ni²⁺. A step gradient of 5% buffer *B* (25 mM Tris–HCl pH 7.5, 500 mM NaCl, 500 mM imidazole) was used to remove histidine-rich proteins. A linear concentration gradient of imidazole from 5 to 50% buffer *B* was applied to elute the product, which was then dialyzed against buffer *C* (25 mM Tris–HCl, 250 mM NaCl pH 7.5) at 277 K overnight in the presence of His-tagged TEV protease. The resulting mixture was applied onto the HisTrap column, which bound the cleaved His tag, TEV protease and uncleaved *SmLip*. The *SmLip* sample from which the His tag had been cleaved was present in the flowthrough. Fractions were analyzed using SDS–PAGE and those containing *SmLip* were pooled. The protein was further purified by size-exclusion chromatography using a Superdex 75 26/60 column (GE Healthcare) equilibrated with buffer *C* on an ÄKTApurifier (GE Healthcare). The column had previously been calibrated with the molecular-weight standards blue dextran (>2000 kDa), thyroglobulin (669 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (43 kDa), carbonic anhydrase (29.5 kDa), ribonuclease A (13.7 kDa) and aprotinin (6.5 kDa) (GE Healthcare; data not shown). The protein eluted as one peak of approximate mass 17 kDa, corresponding to a monomer. Fractions containing the protein were pooled and concentrated to 10 mg ml⁻¹ using Amicon Ultra devices (Millipore) for subsequent use. The purity of the protein was confirmed by SDS–PAGE and mass spectrometry (Fingerprint Proteomics Facility, University of Dundee). A theoretical extinction coefficient of 16 960 M⁻¹ cm⁻¹ at 280 nm was used to estimate the protein concentration (*ProtParam*; Gasteiger *et al.*, 2005); the theoretical mass of one subunit was estimated as 16.1 kDa with a calculated isoelectric point of 5.4. The purified protein sample was stored at 277 K until further use.

2.2. Crystallization, data collection and structure determination

Initial crystallization screens were carried out at 293 K by the sitting-drop vapour-diffusion method in 96-well plates. This was achieved using a Phoenix liquid-handling system (Rigaku, Art Robbins Instruments) and the commercially available PEG (Qiagen) and JCSG+ (Molecular Dimensions) screens. Crystallization occurred in two conditions, which were further optimized using the hanging-drop vapour-diffusion

method with drops consisting of 1 μl protein solution at 10 mg ml⁻¹ in 25 mM Tris-HCl pH 7.5, 250 mM NaCl and 1 μl reservoir solution. The two conditions involved reservoirs consisting of 20% polyethylene glycol 3350, 200 mM KI and of 15% polyethylene glycol 3350, 200 mM NaCl. Monoclinic blocks with minimum dimensions of approximately 0.3 mm grew over 2 d and the addition of glycerol to 10% proved to be a suitable cryoprotectant.

Crystals from the iodide-containing condition were characterized first and data set I was measured in-house using a Rigaku MicroMax-007 rotating-anode X-ray generator (Cu K α , $\lambda = 1.541 \text{ \AA}$) coupled to an R-AXIS IV⁺⁺ image-plate detector. A crystal from the second condition was stored in liquid N₂ and subsequently used to measure a high-resolution data set (data set II) on beamline ID29 at the European Synchrotron Radiation Facility (ESRF; Grenoble, France) using an ADSC Q315R detector. All data were indexed and integrated using XDS (Kabsch, 2010) and scaled using SCALA (Evans, 2006) from the CCP4 program suite (Winn *et al.*, 2011).

Data set I was used to solve the structure by SAD methods targeting the iodides present in the crystallization conditions and to acquire a fairly complete model. The sites of potential anomalous scattering ions or atoms were identified using PHENIX (Adams *et al.*, 2010) and experimental phases were calculated using Phaser (McCoy *et al.*, 2007). Density modification was carried out using histogram matching, averaging on the basis of noncrystallographic symmetry (NCS), and model building was carried out using RESOLVE (Terwilliger, 2003). NCS restraints were employed in the initial refinement calculations, which were performed using REFMAC5 (Murshudov *et al.*, 2011). Inspection of the model and the fit to electron-density and difference density maps was carried out in Coot (Emsley *et al.*, 2010). The analysis then switched to the high-resolution synchrotron data set II when it became available and this was used to complete the refinement. MolProbity (Chen *et al.*, 2010) was used to investigate model geometry in combination with the validation tools provided in Coot. Analyses of surface areas and interactions were made using the PISA (Krissinel & Henrick, 2007) web service and secondary-structure analysis was performed using DSSP (Kabsch & Sander, 1983). Crystallographic statistics are summarized in Table 1.

2.3. Bacterial two-hybrid analyses

For generation of the plasmid pSC072, the gene fragment encoding SmLip amino acids 27–176 was PCR-amplified using primers 5'-TATAgcatgcGTAAAGAGGAGGCTGCATGTC-TTCCGCCAAAAGC-3' and 5'-TATActagaGAGTCGACCTTTTTTACGGGGC-3' and cloned into the vector pUT18 (Karimova *et al.*, 2001) using SphI and XbaI restriction sites. The restriction sites are shown in lower case. For generation of another plasmid, pSC080, the same gene fragment was PCR-amplified using primers 5'-TATAggatccAATGTCTTC-CGCCAAAAGCG-3' and 5'-TATAggtaccAATGATGACG-ACCCCTATCGC-3' and cloned into vector pT25 (Karimova

Table 1
Crystallographic statistics.

Values in parentheses are for the highest resolution shell.

	Data set I	Data set II
Space group	C2	C2
Wavelength (Å)	1.5418	1.007
Unit-cell parameters (Å, °)	$a = 139.7, b = 77.6,$ $c = 54.3, \beta = 98.4$	$a = 139.7, b = 77.8,$ $c = 54.5, \beta = 98.3$
Resolution (Å)	19.7–2.35 (2.48–2.35)	39.8–1.92 (2.02–1.92)
No. of reflections recorded	94650 (11192)	280802 (38597)
Unique reflections	23495 (3109)	43478 (6188)
Completeness (%)	98.3 (90.0)	98.2 (95.8)
Multiplicity	4.0 (3.6)	6.5 (6.2)
$\langle I/\sigma(I) \rangle$	30.9 (6.6)	20.6 (3.7)
Anomalous completeness (%)	95.7 (84.5)	—
Anomalous multiplicity	2.0 (1.8)	—
Wilson B (Å ²)	47.7	32.7
No. of residues/waters	—	541/336
$R_{\text{merge}}^{\dagger}$ (%)	2.6 (17.1)	4.9 (44.2)
$R_{\text{work}}^{\ddagger}$ (%)	—	22.0
R_{free}^{\S} (%)	—	29.2
Average B factors (Å ²)		
Chain A	—	36.1
Chain B	—	41.6
Chain C	—	49.5
Chain D	—	59.8
Waters	—	46.2
Na ⁺	—	37.8
Ethylene glycol	—	60.2
Cruickshank DPI [¶] (Å)	—	0.2
Ramachandran plot		
Most favoured	—	516 residues
Additional allowed	—	21 residues
Outliers	—	Molecule D: Phe97, Asp129; molecule B: Pro142, Ser154
R.m.s.d. on ideal values ^{††}		
Bond lengths (Å)	—	0.01
Bond angles (°)	—	1.42

[†] $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the i th measurement of reflection hkl and $\langle I(hkl) \rangle$ is the mean value of $I_i(hkl)$ for all i measurements. [‡] $R_{\text{work}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$, where F_{obs} is the observed structure factor and F_{calc} is the calculated structure factor. [§] R_{free} is the same as R_{cryst} except calculated with a subset (5%) of data that were excluded from the refinement calculations. [¶] Cruickshank (1999). ^{††} Engh & Huber (1991).

et al., 1998) using BamHI and KpnI restriction sites (again shown in lower case). Bacterial two-hybrid analyses were performed following established protocols (Karimova *et al.*, 1998, 2000). *E. coli* BTH101 was transformed with pSC072 (or pUT18 control) and pSC080 (or pT25 control) and the colour of the resulting transformants was scored on MacConkey media with 0.2% maltose (with a positive result being red). For quantitative measurement of the interaction, β -galactosidase assays were performed as described by Murdoch *et al.* (2011) on double-transformed BTH101 grown at 303 K in Luria–Bertani broth (LB) and permeabilized with toluene. Replicate assays were performed on independent transformants.

2.4. Cellular localization of Lip

Wild-type *S. marcescens* strain Db10 and the *lip* mutant SJC10 (Murdoch *et al.*, 2011) were grown for 8 h at 303 K

in LB. Subcellular fractionation was performed following an established method (Hatzixanthis *et al.*, 2003). In brief, following isolation of clean supernatant by centrifugation, washed cell pellets were resuspended in 50 mM Tris–HCl pH 7.5, 40% (*w/v*) sucrose at 10 ml per gram of cells. EDTA was then added to 5 mM (final concentration) and lysozyme was added to 0.6 mg ml⁻¹ before incubation at 310 K for 30 min. Sphaeroplasts were harvested by centrifugation and taken up in an equivalent volume of 50 mM Tris–HCl pH 7.5 before French pressure treatment. Following ultracentrifugation of the resultant crude extract, the isolated membranes were again taken up in an equivalent volume of 50 mM Tris–HCl pH 7.5. This protocol ensured that equivalent proportions of each cell fraction were assayed. 4 µl of each fraction was mixed with SDS sample buffer (100 mM Tris–HCl pH 6.8, 3.2% SDS, 3.2 mM EDTA, 16% glycerol, 0.2 mg ml⁻¹ Bromophenol blue, 2.5% β-mercaptoethanol) and separated by 15% SDS–PAGE prior to anti-Lip immunoblotting. Whole-cell samples comparing wild-type *versus* SJC10 were prepared by resuspending cells from 100 ml culture in 100 µl SDS sample buffer and boiling for 5 min prior to loading 6 µl onto the gel. Following SDS–PAGE, proteins were electroblotted onto polyvinylidene fluoride membrane (Millipore). *SmLip* was detected by hybridization of the primary antibody polyclonal rabbit anti-Lip (1:4000) followed by the secondary antibody HRP-conjugated goat anti-rabbit (Thermo; 1:10000) and the use of an enhanced chemiluminescent detection kit (Millipore).

3. Results and discussion

3.1. Structure determination

Full-length *SmLip* consists of 176 residues. A truncated version of *SmLip* consisting of an N-terminal hexahistidine tag plus a TEV protease recognition site followed by residues Ala30–Asp176 was obtained in recombinant form and purified. The N-terminal 29 amino acids, which include the lipidation signal peptide and the first four residues of the mature protein, have been omitted. This sample gave monoclinic crystals. The asymmetric unit consists of four polypeptide chains, labelled *A–D*, with an estimated solvent content of 45% and a V_M of 2.27 Å³ Da⁻¹.

Medium-resolution diffraction data were recorded in-house and the anomalous scattering information was used in a SAD approach to phasing. 13 potential iodide positions were identified and produced a figure of merit of 0.43 to 2.35 Å resolution. Subsequently, 12 of these positions were confirmed by refinement with this data set. The initial model constructed in *RESOLVE* consisted of 293 residues, with a correlation coefficient of 0.55 and R_{work} and R_{free} values of 46% and 49%, respectively. The first round of model building in *Coot* extended this to 467 residues, with a correlation coefficient of 0.72 and R_{work} and R_{free} values of 33% and 37%, respectively. At this point the high-resolution synchrotron data (1.92 Å resolution) became available and were used to continue the analysis. The refinement proceeded with the release of NCS

restraints and the incorporation of water molecules, an Na⁺ ion, ethylene glycol and a number of side chains with dual rotamer conformations. This data set was derived from crystals grown in the presence of chloride instead of iodide. However, we did not assign any chloride ions to the structure, noting that typical water molecules occupy the previously identified iodide-binding sites. The refinement was terminated when there were no significant changes in R_{work} and R_{free} and inspection of the difference density map suggested that no further corrections or additions were justified. Several dual rotamers are incorporated into the model. Disorder was evident at several positions, for example the N-terminus, where it was not possible to interpret diffuse and weak electron density. Consequently, several residues are absent from the model. Molecule *A* consists of residues 32–173; molecule *B* of residues 33–142 and 147–176; molecule *C* of residues 34–50, 53–143 and 147–175; and molecule *D* of residues 33–50 and 55–175. The geometry of the model is acceptable (Table 1).

3.2. Self-association and localization *in vivo*

Previous work on SciN, the Lip homologue from entero-aggregative *E. coli*, showed that the protein is localized in the outer membrane, facing the periplasm (Aschtgen *et al.*, 2008). Examination of the amino-acid sequence of the N-terminus of *SmLip* predicts that this is also an outer-membrane lipoprotein. The *LipoP* 1.0 algorithm (Juncker *et al.*, 2003) predicts that *SmLip* has a lipoprotein signal peptide and that signal peptidase II cleavage occurs between Gly25 and Cys26, with the cysteine subsequently being lipidated. Additionally, the residue at the +2 position following cleavage is Met27 (*i.e.* it is not an aspartate, which directs retention in the inner

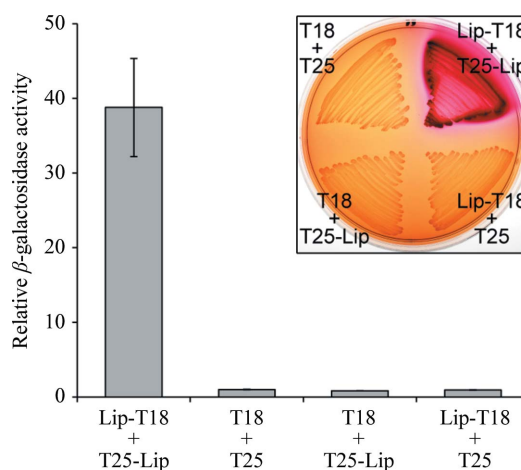


Figure 1

Detection of Lip–Lip self-interaction. The bacterial two-hybrid system was used to detect an *in vivo* interaction between Lip (minus signal peptide) fused to T25 (pSC080) and Lip (minus signal peptide) fused to T18 (pSC072). The empty vectors pUT18 and pT25 represent negative controls. The graph shows the output from the two-hybrid system detected as β-galactosidase activity expressed relative to the pUT18/pT25 baseline level (the maximal β-galactosidase activity observed for the Lip–Lip interaction corresponded to >5000 Miller units). Bars show mean ± SEM. Inset: colourimetric readout of the two-hybrid assay following growth of *E. coli* BTH101 carrying the above plasmids on MacConkey–maltose agar (red is a positive result).

membrane); therefore, *SmLip* should proceed to the outer membrane *via* the Lol system (Bos *et al.*, 2007).

In order to investigate whether *SmLip* undergoes self-interaction, the bacterial two-hybrid system (Karimova *et al.*, 2000) was utilized in *E. coli*. This assay involves reconstitution of adenylate cyclase activity from two non-interacting cyclase fragments, called T18 and T25, from *Bordetella pertussis*. The presence of cyclic AMP activates the transcription of maltose and lactose catabolic operons by *E. coli*. This can be detected by direct measurement of β -galactosidase activity or by using the observation that bacteria capable of fermenting maltose acidify the medium and thus generate a red colour on MacConkey–maltose indicator plates.

SmLip was introduced as both bait and prey by encoding on plasmids pUT18 and pT25, and a strong positive result was observed (Fig. 1). Mature *SmLip* (lacking the N-terminal signal peptide) was used for this experiment, firstly to correspond to the form of *SmLip* for which the structure was solved and secondly to ensure that both partners were localized together in the cytoplasm after fusion with T18 or T25. This positive result indicates that Lip does indeed self-associate within the cell and that neither localization in the outer membrane nor other components of the type VI secretion apparatus are required for self-interaction. We note, however, that this system is unable to distinguish between dimerization or higher order oligomerization.

As a control for any propensity of *SmLip* to form non-specific interactions, in addition to the lack of interaction with the T18 and T25 proteins demonstrated in Fig. 1 we tested whether *SmLip* gave a positive bacterial two-hybrid result with several cytoplasmic components of the T6SS (with which, as a periplasmic protein, it should not interact). *SmLip* gave a negative result (indistinguishable from the T25 + T18 negative control) when tested against the proteins VipB, TssK and TssL (data not shown).

In order to confirm the localization of the native Lip protein in *S. marcescens*, we utilized an anti-Lip polyclonal antibody to probe each of the major cellular fractions in this organism. As shown in Fig. 2, native *SmLip* is found exclusively in the membrane fraction, confirming the predicted localization of the protein and the functionality of the signal peptide.

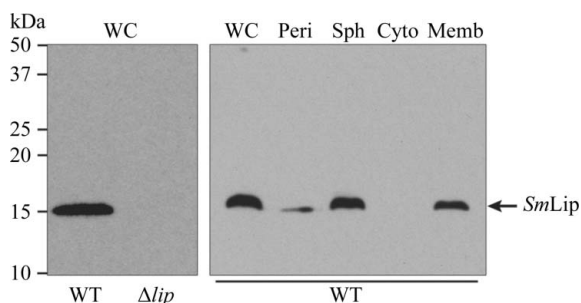


Figure 2
Cellular localization of native *SmLip* in *S. marcescens*. Anti-Lip immunoblot of whole cells or cellular fractions prepared from wild-type *S. marcescens* strain Db10 or the Δlip mutant SJ10 (WC, whole cell; Peri, periplasm; Sph, sphaeroplast; Cyto, cytoplasm; Memb, membranes). The predicted size of mature *SmLip* is 16 kDa.

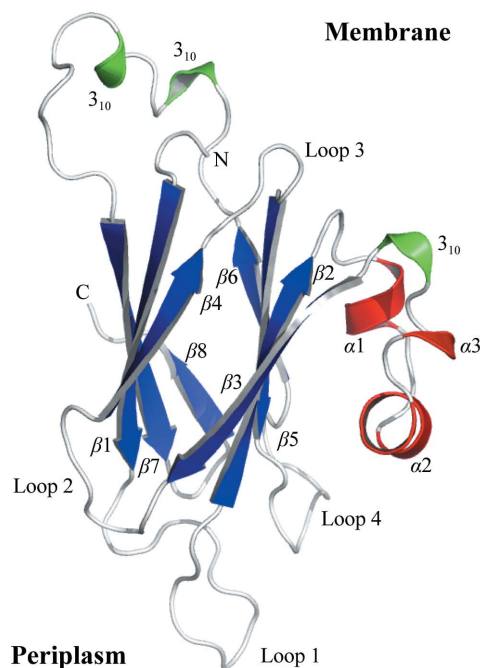


Figure 3
The secondary structure and fold of *SmLip*. β -Strands are shown as blue arrows and α -helices and 3_{10} -turns as red and green ribbons, respectively. The N- and C-terminal residues are labelled and the orientation of the protein with respect to the outer membrane and periplasm is suggested.

3.3. Overall structure

The *SmLip* polypeptide can be classified as a new member of the transthyretin-like superfamily and a detailed comparison will be given below. The protein displays a compact globular structure dominated by an eight-stranded β -sandwich (Fig. 3; Supplementary Fig. S1¹). The order of the strands is 8–7–1–4 and 6–5–2–3. There are three short α -helical segments and three 3_{10} -helix turns. The four *SmLip* polypeptide chains in the asymmetric unit are similar, with the root-mean-square (r.m.s.) deviations between superimposed C^α atoms ranging from 1.3 Å (monomers A and B) to 0.8 Å (monomers A and D) with an average value of 0.95 Å.

Although a set of core conserved proteins are encoded by the T6SS gene clusters in different Gram-negative bacteria (data not shown), there is a large degree of variation in the amino-acid sequences of these proteins. Lip and its orthologues, for example, are relatively poorly conserved. Excluding the signal peptide and lipobox motif (Fig. 4), *SmLip* shares only about 20% sequence identity with SciN, the homologue from enteroaggregative *E. coli*. This increases to near 40% in comparison with the homologue from the *P. aeruginosa* HSI-1 T6SS. Sequence conservation is noted in loop 1, near $\alpha 1$ and $\alpha 2$, in loop 2 and in the loop 4– $\beta 6$ region (Fig. 4).

An alignment of *SmLip* with eight orthologues (Supplementary Fig. S2¹) reinforces the observation of a low level of sequence identity for this protein. Excluding two residues

¹ Supplementary material has been deposited in the IUCr electronic archive (Reference: MN5005). Services for accessing this material are described at the back of the journal.

in the lipobox motif, only six residues are strictly conserved: Asn48, Leu99-*X*-Pro101-Gly102, Gly120 and Ala124. All six residues appear to contribute to the conformation of specific parts of the fold (data not shown). The side chain of Asn48 accepts a hydrogen bond from the main-chain amide of Gln126, helping to define the conformation of loop 4. The Leu99-*X*-Pro101-Gly102 segment defines the structure of the turn after $\beta 3$ leading into loop 2. Gly120 and Ala124 occur in $\beta 5$ and contribute hydrogen bonds to form interactions with $\beta 2$ and $\beta 6$ on either side. An increase in size of the side chain at either of these positions would be likely to be disruptive to the formation of this β -sheet, which forms one side of the structure. There is no obvious hydrophobic, basic or acidic surface feature on *SmLip* that is likely to be conserved within the Lip proteins since the few conserved residues are mainly buried.

The information provided in §3.2 identifies that the N-terminus of the structure is placed close to the outer membrane, hence the assignment of the orientation of *SmLip* with respect to the outer membrane (Fig. 3). By extension, we note that the areas of *SmLip* in which sequence conservation is observed mainly appear to contribute to stabilizing parts of the structure that jut out into the periplasm. They may therefore serve to define the structure of parts of Lip that are responsible for interaction with other molecules in the periplasm.

3.4. The tetramer is likely to be a crystallographic artefact

Gel-filtration data acquired during purification indicated that *SmLip* is a monomer in solution (data not shown). In contrast, the bacterial two-hybrid data revealed a propensity for self-interaction and the asymmetric unit is a tetramer displaying 222 point-group symmetry (Fig. 5). The accessible surface area (ASA) of the *SmLip* polypeptide averages out at approximately 8350 Å²; the range is from 8200 Å² for molecule *D* to 8510 Å² for molecule *A*. Each molecule in the asymmetric unit interacts with two of the other three and two types of protein-protein interface are formed between molecules *A*-*B* and *C*-*D* (interface I) and between molecules *A*-*C* and *B*-*D* (interface II). The type I interface, which is larger, covers an area that is approximately 13% of the ASA of the *SmLip* molecule. Such coverage certainly indicates potential for a biologically relevant dimer. This interface is primarily formed by the antiparallel alignment of two $\beta 7$ strands. Three aromatic residues, Phe147, Trp151 and Phe153, contribute van der

Waals interactions to the association and, by virtue of their relative bulk, also to the ASA (data not shown). The type II interface covers about 6.5% of the ASA of a molecule, a level typical of the values observed simply owing to molecular packing in a crystal lattice. This interface is formed by the antiparallel alignment of two $\beta 4$ strands. The areas of *SmLip* involved in forming a tetramer are not conserved in the homologues from *E. coli* or *P. aeruginosa* (Fig. 4) and it is unlikely that such a tetramer is a generic feature of this lipoprotein.

The spatial placement of the N-terminal residues in the asymmetric unit is such that it is unlikely that an oligomeric assembly could form when the protein is anchored in the membrane by the lipidated Cys26 at the N-terminus. The N-termini of molecules *A* and *D* are on the same side of the tetrameric assembly but are opposite to those of molecules *B* and *C*. As explained, there are no direct interactions formed between molecules *A* and *D* or molecules *B* and *C*. That the bacterial two-hybrid experiments reveal a propensity for self-interaction of the truncated protein in the cytoplasm is in one sense consistent with the crystal structure of the truncated version of *SmLip*, which shows a tetrameric assembly containing a plausible dimer. On the other hand, the structure of the tetramer is incompatible with dimeric or tetrameric

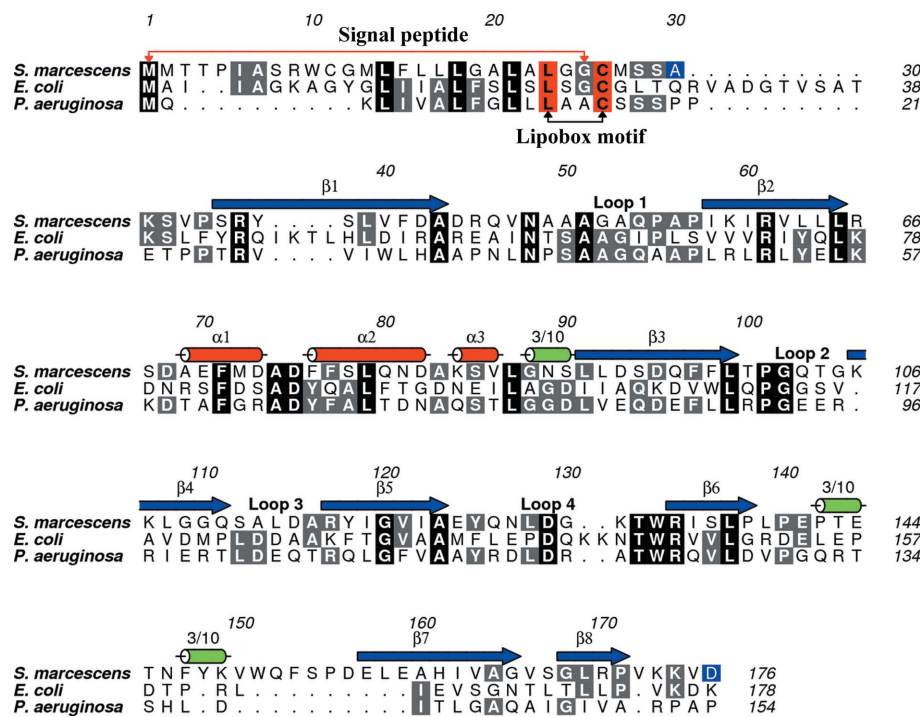


Figure 4 The primary and secondary structure of *SmLip* and sequence alignment with two homologues. *S. marcescens* Lip is aligned with the homologous proteins from enteroaggregative *E. coli* (GenBank CBG37366.1) and *P. aeruginosa* (NCBI Reference Sequence NP_248770.1, PA0080). The secondary structure of *SmLip* is depicted with blue arrows for β -strands and red and green cylinders for α -helices and 3_{10} -helices, respectively. Residues conserved in all three sequences are shown in black boxes and those conserved in only two sequences are shown in grey boxes. The start and finish of the lipobox motif are marked by red boxes; the residues at the start and end of the sequence used in the structure analysis (Ala30-Asp176) are shown in blue boxes. The alignment was generated using *T-Coffee* (Di Tommaso *et al.*, 2011) in the M-Coffee mode and the figure was prepared using *ALINE* (Bond & Schüttelkopf, 2009).

structures if the N-terminus is membrane-bound. These observations may be a result of the different concentrations and experimental conditions used. We suggest that *SmLip* is a membrane-bound monomer but displays a propensity to interact with itself.

A reviewer commented on the possibility that the *SmLip* tetramer might represent an inactive or alternative state of the protein. This is an intriguing suggestion and raises questions about how conversion to an active form might occur and how the T6SS itself is regulated. We have no data to address this issue and further studies would be required to investigate such a possibility.

3.5. Comparisons with structural homologues

A search for structural neighbours in the Protein Data Bank using the *PDBeFold* (<http://pdbe.org/fold>) and *ProFunc* servers (Laskowski *et al.*, 2005) gives a *Z* score of 6.1 with sea bream transthyretin (Eneqvist *et al.*, 2004; PDB entry 1sn0). This matched 84 residues with an r.m.s.d. of 2.7 Å. The β -sheet structures align well (Supplementary Fig. S3). The r.m.s.d. and relatively low *Z* score reflect the low sequence identity shared between the two proteins of approximately 7%. Nevertheless, the structural relationship is clear and *SmLip* can be classed as a new member of the transthyretin-like protein family. The only other member of this protein family is 5-hydroxyisourate hydrolase (EC 3.5.2.17; Hennebray *et al.*, 2006), an enzyme that is found only in prokaryotes, leading to the conclusion that this represents an example of divergent evolution (Hennebray, 2009). The sequence identity shared between this hydrolase and *SmLip* is only 6%, but the similarity in fold is evident (data not shown). We carried out further comparisons seeking to inform on Lip function.

Transthyretin binds the hormone thyroxine, self-interacts to form a tetramer and also forms a complex with retinol-binding

protein (Blake *et al.*, 1978; Wojtczak *et al.*, 1992; Monaco *et al.*, 1995; Zanotti *et al.*, 2008). In common with transthyretin, *SmLip* forms a tetrameric assembly. However, the *SmLip* oligomer is distinct and an overlay of one *SmLip* polypeptide with a subunit from transthyretin does not produce an overlap of any of the other polypeptides (data not shown).

Transthyretin forms a dimer by antiparallel self-association of the β_6 and β_8 strands, creating a curved eight-stranded β -sheet (Blake *et al.*, 1978). The binding of the hormone thyroxine occurs at the tetramer interface created by the convex surfaces of two of these eight-stranded β -sheets as the protein assembles as a dimer of dimers. The thyroxine-binding residues in transthyretin are not conserved in *SmLip* and an overlay of an *SmLip* polypeptide and transthyretin subunit places the ligand-binding site on the surface of the former (Supplementary Fig. S3). It is unlikely that *SmLip* acts to bind hydrophobic ligands of the type that transthyretin can bind.

Transthyretin associates with retinol-binding protein using residues in three turns: two from one subunit that link β_1 to β_2 and β_4 to β_5 , and one from another subunit that links β_1 to β_2 (Monaco *et al.*, 1995). These parts of the transthyretin structure correspond to loops 1 and 3 of *SmLip*. Loop 1 is directed out from the globular fold into the periplasmic space; it is placed to interact with physiological partners and may represent a binding site for other proteins/molecules.

In a recent study of *Klebsiella pneumoniae* 5-hydroxyisourate hydrolase, the residues important for catalytic function were confirmed as His7, Arg41, His92 and Ser108, which together with Tyr105 form a polar and symmetric active site at a dimer interface (French & Ealick, 2011). A structure-based sequence alignment matches four of these catalytic residues (with the exception being Ser108) to Asp42, Gly105, His161 and Val172, respectively, in *SmLip*. The polypeptides do not overlay in the vicinity of Ser108 (data not shown) and it is unlikely that Lip has any hydrolase activity.

The biological role of *SmLip* or its orthologues in the T6SS has yet to be unambiguously defined. Structural comparisons appear to rule out, rather than assign, a function. The propensity to self-associate using parts of the *SmLip* structure that will be exposed in the periplasm suggests that this protein, exploiting the lipid anchor, helps to bind and position different components of the secretion apparatus at the outer membrane. Future experiments, aided by the structural model, can address this hypothesis.

This work was funded by the European Commission Seventh Framework Programme (FP7/2007-2013), the Aero-path project, the Wellcome Trust (grants 082596 and 083481 and a PhD studentship to GE) and a Royal Society Project Grant and a Royal Society of Edinburgh Personal Research Fellowship (SJC). We acknowledge the contribution of the Pathogen Sequencing Unit at the Wellcome Trust Sanger Institute, Hinxton, England for performing the *S. marcescens* Db11 genome-sequencing project.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.

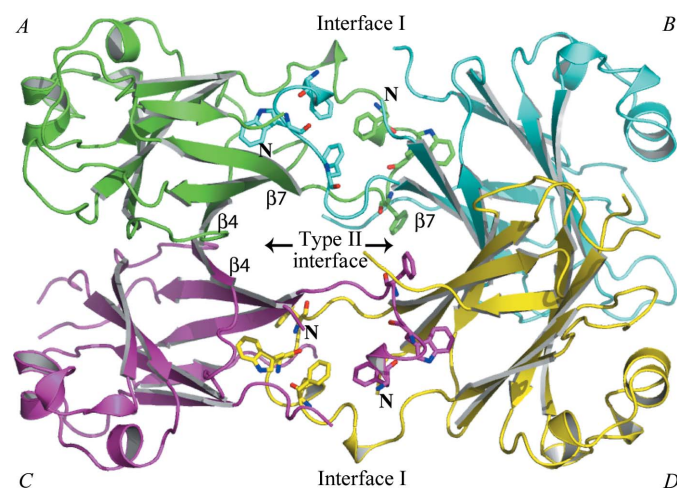


Figure 5

The asymmetric unit. The four molecules that constitute the asymmetric unit are shown in different colours using the secondary-structure assignment given in Fig. 1 and labelled. The two types of protein–protein interface are labelled, as are the β_4 and β_7 strands. Residues that contribute significantly to the type I interface (Phe147, Trp151 and Phe153) are depicted as sticks.

- Aschtgen, M. S., Bernard, C. S., De Bentzmann, S., Llobès, R. & Cascales, E. (2008). *J. Bacteriol.* **190**, 7523–7531.
- Blake, C. C., Geisow, M. J., Oatley, S. J., Rérat, B. & Rérat, C. (1978). *J. Mol. Biol.* **121**, 339–356.
- Blondel, C. J., Yang, H.-J., Castro, B., Chiang, S., Toro, C. S., Zaldívar, M., Contreras, I., Andrews-Polymenis, H. L. & Santiviago, C. A. (2010). *PLoS One*, **5**, e11724.
- Bond, C. S. & Schüttelkopf, A. W. (2009). *Acta Cryst.* **D65**, 510–512.
- Bönemann, G., Pietrosiuk, A. & Mogk, A. (2010). *Mol. Microbiol.* **76**, 815–821.
- Bos, M. P., Robert, V. & Tommassen, J. (2007). *Annu. Rev. Microbiol.* **61**, 191–214.
- Cascales, E. (2008). *EMBO Rep.* **9**, 735–741.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Cruickshank, D. W. J. (1999). *Acta Cryst.* **D55**, 583–601.
- Di Tommaso, P., Moretti, S., Xenarios, I., Orobítz, M., Montanyola, A., Chang, J.-M., Taly, J.-F. & Notredame, C. (2011). *Nucleic Acids Res.* **39**, W13–W17.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Eneqvist, T., Lundberg, E., Karlsson, A., Huang, S., Santos, C. R., Power, D. M. & Sauer-Eriksson, A. E. (2004). *J. Biol. Chem.* **279**, 26411–26416.
- Engh, R. A. & Huber, R. (1991). *Acta Cryst.* **A47**, 392–400.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Filloux, A., Hachani, A. & Bleves, S. (2008). *Microbiology*, **154**, 1570–1583.
- French, J. B. & Ealick, S. E. (2011). *Acta Cryst.* **D67**, 671–677.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa: Humana Press.
- Gerlach, R. G. & Hensel, M. (2007). *Int. J. Med. Microbiol.* **297**, 401–415.
- Hatzixanthis, K., Palmer, T. & Sargent, F. (2003). *Mol. Microbiol.* **49**, 1377–1390.
- Hejazi, A. & Falkiner, F. R. (1997). *J. Med. Microbiol.* **46**, 903–912.
- Hennebry, S. C. (2009). *FEBS J.* **276**, 5367–5379.
- Hennebry, S. C., Law, R. H., Richardson, S. J., Buckle, A. M. & Whisstock, J. C. (2006). *J. Mol. Biol.* **359**, 1389–1399.
- Holland, I. B. (2010). *Methods Mol. Biol.* **619**, 1–20.
- Hood, R. D. *et al.* (2010). *Cell Host Microbe*, **7**, 25–37.
- Jani, A. J. & Cotter, P. A. (2010). *Cell Host Microbe*, **8**, 2–6.
- Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H. & Krogh, A. (2003). *Protein Sci.* **12**, 1652–1662.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
- Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 5752–5756.
- Karimova, G., Ullmann, A. & Ladant, D. (2000). *Methods Enzymol.* **328**, 59–73.
- Karimova, G., Ullmann, A. & Ladant, D. (2001). *J. Mol. Microbiol. Biotechnol.* **3**, 73–82.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Laskowski, R. A., Watson, J. D. & Thornton, J. M. (2005). *Nucleic Acids Res.* **33**, W89–W93.
- Liu, H., Coulthurst, S. J., Pritchard, L., Hedley, P. E., Ravensdale, M., Humphris, S., Burr, T., Takle, G., Brurberg, M. B., Birch, P. R., Salmund, G. P. & Toth, I. K. (2008). *PLoS Pathog.* **4**, e1000093.
- MacIntyre, D. L., Miyata, S. T., Kitaoka, M. & Pukatzki, S. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 19520–19524.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Monaco, H. L., Rizzi, M. & Coda, A. (1995). *Science*, **268**, 1039–1041.
- Murdoch, S. L., Trunk, K., English, G., Fritsch, M. J., Pourkarimi, E. & Coulthurst, S. J. (2011). *J. Bacteriol.* **193**, 6057–6069.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Pukatzki, S., McAuley, S. B. & Miyata, S. T. (2009). *Curr. Opin. Microbiol.* **12**, 11–17.
- Sarris, P. F., Skandalis, N., Kokkinidis, M. & Panopoulos, N. J. (2010). *Mol. Plant Pathol.* **11**, 795–804.
- Schwarz, S., Hood, R. D. & Mougous, J. D. (2010). *Trends Microbiol.* **18**, 531–537.
- Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
- Terwilliger, T. C. (2003). *Acta Cryst.* **D59**, 38–44.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Wojtczak, A., Luft, J. & Cody, V. (1992). *J. Biol. Chem.* **267**, 353–357.
- Zanotti, G., Folli, C., Cendron, L., Alfieri, B., Nishida, S. K., Gliubich, F., Pasquato, N., Negro, A. & Berni, R. (2008). *FEBS J.* **275**, 5841–5854.