Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the cytoplasmic domain of TcpE, the inner membrane core protein required for assembly of the *Vibrio cholerae* toxin-coregulated pilus

Subramaniapillai Kolappan and Lisa Craig*

Molecular Biology and Biochemistry Department, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

Correspondence e-mail: licraig@sfu.ca

Received 19 October 2012
Accepted 10 December 2012

PDB Reference: TcpE$^{1-102}$, 4hhx

Type IV pili are long thin surface-displayed polymers of the pilin subunit that are present in a diverse group of bacteria. These multifunctional filaments are critical to virulence for pathogens such as *Vibrio cholerae*, which use them to form microcolonies and to secrete the colonization factor TcpF. The type IV pili are assembled from pilin subunits by a complex inner membrane machinery. The core component of the type IV pilus-assembly platform is an integral inner membrane protein belonging to the GspF superfamily of secretion proteins. These proteins somehow convert chemical energy from ATP hydrolysis by an assembly ATPase on the cytoplasmic side of the inner membrane to mechanical energy for extrusion of the growing pilus filament out of the inner membrane. Most GspF-family inner membrane core proteins are predicted to have N-terminal and central cytoplasmic domains, cyto1 and cyto2, and three transmembrane segments, TM1, TM2 and TM3. Cyto2 and TM3 represent an internal repeat of cyto1 and TM1. Here, the 1.88 Å resolution crystal structure of the cyto1 domain of *V. cholerae* TcpE, which is required for assembly of the toxin-coregulated pilus, is reported. This domain folds as a monomeric six-helix bundle with a positively charged membrane-interaction face at one end and a hydrophobic groove at the other end that may serve as a binding site for partner proteins in the pilus-assembly complex.

1. Introduction

The type IV pili (T4P) are a unique class of pili that have a diverse range of functions and play key roles in bacterial colonization and pathogenesis. Although the functions of T4P vary from one species to another, they include adhesion, self-aggregation to form bacterial microcolonies, twitching motility and DNA uptake in natural transformation. At least some of these functions require dynamic assembly and disassembly or retraction of the pilus. The T4P system is closely related to the type II secretion (T2S) system with respect to amino-acid sequence and structure of the core assembly components (Nunn, 1999; Pugsley, 1993; Sandkvist, 2001; Whitchurch *et al.*, 1991). In fact, some T4P perform secretion themselves, transporting proteins from the periplasm across the outer membrane *via* a secretin channel (Han *et al.*, 2007; Kirn *et al.*, 2003). The T4P are comprised of thousands of copies of the major pilin subunit which form long thin filaments on the bacterial surface, whereas T2S systems are thought to polymerize pseudopilins to form a short periplasmic pseudopilus that acts as a piston to extrude substrate across the outer membrane secretin. Both systems would seem to require rapid

© 2013 International Union of Crystallography
Printed in Singapore – all rights reserved

and processive filament assembly and disassembly, yet these processes are poorly understood.

T4P are assembled by a multicomponent protein apparatus that spans both bacterial membranes. T4P assembly requires ten or more proteins, but only a handful of these are conserved among all T4P systems and have homologs in the T2S system (Sandkvist, 2001; Ayers *et al.*, 2010). These key assembly components are (i) the pilin subunit, which has a hydrophobic N-terminus that serves as a membrane anchor prior to pilus assembly and as a polymerization domain in the intact filament (Craig & Li, 2008; Giltner *et al.*, 2012); (ii) a prepilin peptidase located in the inner membrane that processes the pilin subunits (Kaufman *et al.*, 1991; LaPointe & Taylor, 2000; Lory & Strom, 1997; Zhang *et al.*, 1994); (iii) a hexameric nucleotide-binding protein of the GspE family of secretion ATPases (Planet *et al.*, 2003), which associates with the cytoplasmic side of the inner membrane and hydrolyzes ATP to drive filament assembly (Crowther *et al.*, 2004; Freitag *et al.*, 1995; Nunn, 1999; Taylor *et al.*, 1987); (iv) an integral inner membrane core protein (IMCP) of the GspF secretion family that may link the assembly ATPase on the cytoplasmic side of the inner membrane to the growing filament on its periplasmic side (Chiang *et al.*, 2008; Kaufman *et al.*, 1993; Nunn, 1999; Tønjum *et al.*, 1995; Yamagata *et al.*, 2012) and (v) an outer membrane secretin through which the assembled pilus passes (Bose & Taylor, 2005; Nudleman *et al.*, 2006; Schmidt *et al.*, 2001; Tønjum *et al.*, 1998). Additional essential proteins that are less conserved among the T4P systems include inner membrane accessory proteins that form an inner membrane assembly platform together with the core protein and assembly ATPase (Ayers *et al.*, 2009; Tripathi & Taylor, 2007; Crowther *et al.*, 2004). Importantly, many T4P utilize a 'retraction ATPase' to disassemble pili (Anantha *et al.*, 1998; Chiang *et al.*, 2005; Merz *et al.*, 2000; Misic *et al.*, 2010; Wolfgang *et al.*, 1998). While a considerable amount is known about the T4P structure and the ATPase motors of the T4P and T2S systems, much less is known about the role of the IMCP and about the molecular events that link the conformational change in cytoplasmic ATPase motor to growth of the filament out of the inner membrane and through the periplasm.

The T4P IMCP and other GspF secretion proteins display an internal repeat, with two tandem GspII_F motifs each corresponding to a predicted cytoplasmic domain followed by a transmembrane (TM) segment (Peabody *et al.*, 2003). An intervening TM segment yields a 3-TM topology, which has been confirmed experimentally for several of the secretion IMCPs using reporter-enzyme fusions (Arts *et al.*, 2007; Thomas *et al.*, 1997; Abendroth *et al.*, 2009). A 4-TM topology has been described for the enteropathogenic *Escherichia coli* (EPEC) IMCP BfpE (Blank & Donnenberg, 2001), but this IMCP may be an outlier. Crystal structures have been determined for the N-terminal cytoplasmic domain of the GspF orthologs EpsF from the *Vibrio cholerae* T2S system (Abendroth *et al.*, 2009) and PilC from the *Thermus thermophilus* T4P system (Karuppiah *et al.*, 2010). Attempts by Hol and coworkers to crystallize the entire N-terminal cytoplasmic domain of *V. cholerae* EpsF (residues 1–171) and its homologs were unsuccessful, but a crystal structure was determined for a fragment spanning residues 56–171 ($EpsF^{56–171}$). In the case of *T. thermophilus*, the cytoplasmic domain of PilC (residues 1–168) was partially degraded during crystallization, producing a structure spanning residues 53–168 ($PilC^{53–168}$). These results suggest that the most N-terminal ∼50 residues are disordered, at least in the purified proteins. The $EpsF^{56–171}$ and $PilC^{53–168}$ structures are very similar six-helix bundles. Importantly, $EpsF^{56–171}$ has an extended C-terminal α-helix followed immediately by TM1 in the native protein and thus defines the membrane orientation of this domain. Both proteins form dimers in the crystal lattice, and $PilC^{53–168}$ also forms a dimer in solution. The N-terminal domain of the EPEC IMCP BfpE interacts with itself, as shown by yeast two-hybrid studies (Crowther *et al.*, 2004), suggesting that the IMCPs form multimers *via* their cytoplasmic domains. This idea was supported by a low-resolution electron-microscopy reconstruction of the *Neisseria meningitidis* T4P IMCP PilG, which forms a tetramer when solubilized in detergent (Collins *et al.*, 2007).

The *V. cholerae* toxin-coregulated pili (TCP) belong to a subclass of T4P, type IVb, which includes the bundle-forming pilus from EPEC and CFA/III and longus from enterotoxigenic *E. coli*. The type IVb pilins are longer than the type IVa pilins and have a distinct fold (Craig & Li, 2008), and their biogenesis apparatus has fewer components, all of which are encoded in a single gene cluster. This contrasts with the more complex type IVa pilus systems, including those of *N. meningitidis* and *T. thermophilus*, which require as many as 40 genes distributed throughout the genome. Thus, the type IVb pili may represent more tractable systems in which to study pilus assembly. In *V. cholerae*, the IMCP TcpE, together with the inner membrane accessory proteins TcpR and TcpD and the assembly ATPase TcpT, form the pilus assembly platform. TcpR and TcpD are each predicted to have a single TM segment with either a cytoplasmic N-terminal domain (TcpR) or a periplasmic C-terminal domain (TcpD). TcpR interacts with the N-terminal ∼100 residues of the ATPase TcpT and localizes it to the inner membrane (Tripathi & Taylor, 2007). An analogous interaction is observed for the *V. cholerae* T2S system in the crystal structure of a complex between the cytoplasmic N-terminal domains of the inner membrane accessory protein EpsL and the assembly ATPase EpsE (Abendroth *et al.*, 2004). Little is known about the function of the *V. cholerae* IMCP TcpE or any of its secretion-family homologs. Here, we report the X-ray crystal structure of the N-terminal cytoplasmic domain of *V. cholerae* TcpE from the type IVb pilus system and discuss its implications for pilus assembly.

2. Materials and methods

2.1. Cloning, protein expression and purification

The genes encoding cyto1-TcpE (residues 1–116) and cyto2-TcpE (residues 187–303) were PCR-amplified from *V. cholerae* strain O395 genomic DNA using forward primers

5′-AAGGTTCCATATGATGAAAATTATCTCCAAGAAG-TATAGGC-3′ and 5′-AAGGTTCCATATGACAGGGAA-TTTCAGAGATGGTTTTTTAG-3′ and reverse primers 5′-CGCGGATCCTTAGCTAGGCGTGATCATAGATGAG-ATAGC-3′ and 5′-CGCGGATCCTTAAATTTTTTTATG-TAGATTTATATTGGCTT-3′. PCR products were digested with *Nde*I and *Bam*HI and ligated into expression vector pET-15b, which encodes an N-terminal hexahistidine tag and a thrombin cleavage site. Plasmids were transformed into *E. coli* RosettaBlue (DE3) (Novagen) and cells were grown in Luria–Bertani (LB) broth containing 100 µg ml^{-1} ampicillin at 310 K to an OD_{600} of ∼0.4. Protein expression was induced with 0.4 m*M* isopropyl β-D-1-thiogalactopyranoside and cells were grown for a further 4 h at 310 K before harvesting by centrifugation (5000*g*, 30 min, 277 K). Cell pellets from 6 l culture were frozen in liquid nitrogen, thawed on ice, suspended in 100 ml lysis buffer (50 m*M* Na_2HPO_4/NaH_2PO_4 pH 7.4, 50 m*M* NaCl, 1 m*M* EDTA, 1 mg ml^{-1} lysozyme) and gently stirred at room temperature for 1 h. Cells were lysed by sonication and cell debris was removed by centrifugation (40 000*g*, 60 min, 277 K). The supernatant was filtered through a 0.4 µm polyethersulfone membrane and loaded onto an Ni–NTA column pre-equilibrated with buffer *A* (50 m*M* Na_2HPO_4/NaH_2PO_4 pH 7.4, 30 m*M* imidazole, 50 m*M* NaCl, 1 m*M* EDTA); it was then eluted with 240 m*M* imidazole pH 7.4 in buffer *A* followed by overnight dialysis against buffer *B* (20 m*M* Tris–HCl pH 7.4, 50 m*M* NaCl, 1 m*M* EDTA). The His tag was removed from a portion of the cyto1-TcpE by thrombin treatment followed by size-exclusion chromatography in buffer *B*. Fractions were pooled and concentrated to 15 mg ml^{-1} using a stirred-cell concentrator (Millipore) with a 3000 Da molecular mass cutoff membrane. The purity of the cyto1-TcpE was estimated to be >95% for both the untagged and His-tagged proteins by SDS–PAGE analysis. The proteins were flash-frozen in liquid nitrogen in 100 µl aliquots and stored at 193 K prior to crystallization. Selenomethionine-substituted cyto1-TcpE (SeMet-cyto1-TcpE) was prepared by growing cells in minimal medium supplemented with selenomethionine instead of methionine, plus the other 19 amino acids (Doublié, 1997), and was purified using the same procedure as for native cyto1-TcpE, including removal of the His tag.

Glu82 substitutions were introduced into cyto1-TcpE by QuikChange mutagenesis of the *tcpE* gene in pET-15b using the primers 5′-ATGATTAATGTTGCAGCAAAC-TCAGGTAAGATTT-3′ and 5′-AAATCTTACCTGAGTTT-GCTGCAACATTAATCAT-3′ for Glu82Ala and 5′-ATG-ATTAATGTTGCAAAAAACTCAGGTAAGATTT-3′ and 5′-AAATCTTACCTGAGTTTTTTGCAACATTAATCAT-3′ for Glu82Arg. Expression of the cyto1-TcpE^{E82A} and cyto1-TcpE^{E82R} variants under the same conditions as the native protein (310 K) was unsuccessful. These proteins did overexpress well at lower temperatures (293 and 303 K); however, they were insoluble. Cell fractions containing the insoluble protein were resuspended in lysis buffer containing 0.1% Triton X-100 and 0.1% Tween and incubated at room temperature for 1 h. The resuspensions were centrifuged for 30 min at 10 000*g*. The pellets were then resuspended in lysis buffer containing 8 *M* urea and incubated overnight at 277 K. The denatured protein was loaded onto an Ni–NTA column and washed with buffer *A* containing 8 *M* urea followed by buffer *A* alone. Protein was eluted with buffer *A* containing 300 m*M* imidazole pH 7.4. Most of the protein remained bound to the column and the small amount of protein that eluted from the column was insoluble.

Table 1
Crystallographic data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell (1.99–1.93 Å).

	SeMet-cyto1-TcpE	SeMet-cyto1-TcpE	
Data collection			
Beamline	SSRL 9-2	SSRL 9-2	
Space group	$C222_1$	$C222_1$	
Unit-cell parameters (Å, °)	a = 34.8, b = 73.6, c = 95.9, $\alpha = \beta = \gamma$ = 90.0	a = 34.7, b = 73.7, c = 95.7, $\alpha = \beta = \gamma$ = 90.0	
Resolution (Å)	1.88	2.7	2.7
Wavelength (Å)	0.9116	0.9611	0.9791
Completeness (%)	99.5 (96.2)	99.9 (99.9)	99.9 (99.9)
Observed reflections	85767	30526	30553
Unique reflections	10378	3599	3598
R_{merge}† (%)	6.2 (71.8)	5.2 (9.0)	5.6 (9.6)
$\langle I/\sigma(I)\rangle$	18.0 (2.3)	10.5 (7.6)	9.7 (7.0)
Wilson B value (Å^2)	39.06	44.6	45.6
Mosaicity (°)	0.20	0.5	0.5
Refinement statistics			
Resolution limits (Å)	20.0–1.88		
Molecules per asymmetric unit	1		
No. of reflections used	9890		
R_{cryst}‡ (%)	21.8		
R_{free}§ (%)	24.7		
No. of atoms			
Protein	831		
Ligand	5		
Water	25		
Average B factor (Å^2)			
Protein	38.5		
Sulfate ions	37.2		
Water O atoms	40.3		
R.m.s.d., bond lengths (Å)	0.006		
R.m.s.d., bond angles (°)	1.0		
Ramachandran plot statistics (%)			
Favored (%)	98.0		
Allowed (%)	2.0		

† R_{merge} is the unweighted R value on I between symmetry mates. ‡ $R_{cryst} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$. § R_{free} is the cross-validation R factor for 5% of reflections against which the model was not refined.

2.2. Protein crystallization

Initial crystallization conditions for His-tagged and untagged cyto1-TcpE were determined by mixing 2 µl cyto1-TcpE (15 mg ml^{-1} in buffer *B*) with 2 µl reservoir solution from commercially available screens using the hanging-drop vapor-diffusion method. Crystals were grown at 293 K. Thin plate-shaped crystals of untagged cyto1-TcpE were obtained in 100 m*M* Tris–HCl pH 7.4, 1.6 *M* ammonium sulfate. SeMet-cyto1-TcpE crystals were grown in 100 m*M* Tris–HCl pH 7.4, 1.55 *M* ammonium sulfate. The crystals were flash-cooled in liquid nitrogen with cryoprotectant [25%(*v*/*v*) glycerol in reservoir buffer] and stored in liquid nitrogen.

2.3. Data collection and structure determination

X-ray diffraction data for SeMet-cyto1-TcpE were collected at 100 K on beamline 9-2 of Stanford Synchrotron Radiation Lightsource (SSRL) using the *Blu-Ice* software, a graphical user interface for crystallographic data collection (Gonzalez *et al.*, 2008). Data sets were collected at two wavelengths corresponding to the high-energy remote and the inflection point, as determined by an X-ray fluorescence scan, for multi-wavelength anomalous diffraction (MAD) phasing. The data sets were processed and scaled to 2.7 Å resolution using the *CCP*4 suite (Winn *et al.*, 2011) programs *iMosflm* (Battye *et al.*, 2011) and *SCALA* (Leslie, 1992), respectively. Phases were determined using *SOLVE* (Terwilliger & Berendzen, 1999) and *RESOLVE* (Terwilliger, 2000). Next, a diffraction data set was collected from an SeMet-cyto1-TcpE crystal that diffracted to better than 2 Å resolution. This data set was processed and scaled with the *XDS* suite (Kabsch, 2010) to 1.88 Å resolution and phase extension was carried out using the *CCP*4 program *DM*, which yielded an interpretable map. The model was built and refined using *Coot* (Emsley & Cowtan, 2004) and *REFMAC* (Murshudov *et al.*, 2011) and was validated using *PROCHECK* (Laskowski *et al.*, 1993) and *MolProbity* (Chen *et al.*, 2010).

3. Results and discussion

3.1. Expression and crystallization of Cyto1-TcpE

V. cholerae TcpE is predicted to be a polytopic protein with three transmembrane helices (TM1, TM2 and TM3), an N-terminal cytoplasmic domain (cyto1) and a central cytoplasmic domain (cyto2). A short ∼30-amino-acid loop on the periplasmic side of TcpE connects TM1 and TM2, and TM3 is followed by an approximately seven-residue periplasmic C-terminal segment (Fig. 1). Cyto1-TcpE and its corresponding domains in other type IVb pilus IMCPs are relatively short (89–118 amino acids) compared with those of the T2S and type IVa pilus systems (141–170 amino acids). Amino-acid sequence alignment indicates that the type IVb IMCPs lack the protease-sensitive disordered N-terminal segment present in the type IVa and T2S IMCPs (Fig. 1). Constructs were designed to express both of the cytoplasmic domains, cyto1-TcpE (residues 1–116) and cyto2-TcpE (residues 187–303), separately in *E. coli*. Both domains expressed well, but only cyto1-TcpE was soluble. Attempts were made to obtain soluble cyto2-TcpE, including varying the expression strain, medium, induction time, expression temperature and composition of the lysis buffer as well as refolding from inclusion bodies and co-expression with cyto1-TcpE, all of which were unsuccessful. Cyto1-TcpE was purified using metal-affinity and size-exclusion chromatography. For a portion of the purified cyto1-TcpE, the His tag was removed by thrombin digestion and size-exclusion chromatography. Untagged cyto1-TcpE eluted from the size-exclusion column with a molecular mass of ∼14 kDa, which is consistent with a monomer (calculated mass 13.2 kDa). Crystallization trials were performed using both His-tagged and untagged protein. Thin plate-like crystals were obtained of untagged cyto1-TcpE that diffracted to ∼5 Å resolution. The limited sequence identity of cyto1-TcpE to GspF-family members of known structure (19.8% to *V. cholerae* EpsF and 14.3% to *T. thermophilus* PilC overall) meant that molecular replacement was unlikely to provide a structure solution, so SeMet-substituted protein was prepared in order to solve the structure by MAD phasing. Crystallization conditions for SeMet-cyto1-TcpE were optimized from those of the native cyto1-TcpE crystals, resulting in large plates. A dual-wavelength diffraction data set was collected from an SeMet-cyto1-TcpE crystal in space group $C222_1$ and its structure was solved to 2.7 Å resolution using MAD methods (Table 1). Subsequently, a 1.88 Å resolution structure was solved from a new SeMet-cyto1-TcpE crystal in the same space group by phase extension.

3.2. Cyto1-TcpE crystal structure

Residues 1–102 of the 116-amino-acid cyto1-TcpE construct were resolved in the crystal structure, which is referred to here as $TcpE^{1–102}$. $TcpE^{1–102}$ forms a six-helix bundle, with the antiparallel α-helices loosely arranged in a ring or cylinder that measures ∼30 Å in height and ∼33 Å in diameter (Figs. 2*a* and 2*b*). Each helix is tilted relative to the cylinder axis and has 4–5 turns, with the exception of the 1.5-turn α4. The N- and C-terminal ends of cyto1-TcpE are both located on the top of the cylinder, with the N-terminal segment extending across the top of the cylinder prior to entry into α1. In full-length TcpE TM1 is predicted to immediately follow α6 (Fig. 1), which suggests that the top of the cylinder is closely apposed to the cytoplasmic face of the inner membrane (Fig. 2*a*). Consistent with this prediction, there is a high concentration of lysines and arginines on this face that could interact with the negatively charged phosphates of the membrane phospholipids (Figs. 2*a*, 2*b* and 3*c*). The core of the cylinder is comprised of hydrophobic side chains, many of which are conserved in the GspF family proteins. The bottom of the cylinder has a pronounced groove lined with hydrophobic side chains from the cylinder core (Figs. 2*c* and 2*d*). On one lip of this groove lie three highly conserved residues, Glu82, Ser84 and Gly85, which are located at the C-terminal end of α5 (Figs. 1*b* and 2). The side-chain O atoms of Glu82 hydrogen-bond to Thr63 O^{γ} and N and Leu62 N at the N-terminal end of the adjacent α4 (Fig. 2*a*). The side-chain hydroxyl of Ser84, the terminal residue of α5, hydrogen-bonds to the backbone O atoms of Val80 and Ala81 on the same α-helix. Gly85 lies at the point of an irregular turn between α5 and α6 that protrudes from the bottom of the cylinder.

3.3. Comparison of the crystal structure of cyto1-TcpE with those of $EpsF^{56–171}$ and $PilC^{53–168}$ and the role of Glu82

Despite relatively low sequence similarity between TcpE and EpsF or PilC in their N-terminal cytoplasmic domains (14.3% identity and 40.2% homology between cyto1-TcpE and cyto1-EspF; 15.2% identity and 44.6% homology between cyto1-TcpE and $PilC^{53–168}$), their structures are conserved: $TcpE^{1–102}$ superimposes on $EpsF^{56–171}$ (PDB entry 3c1q;

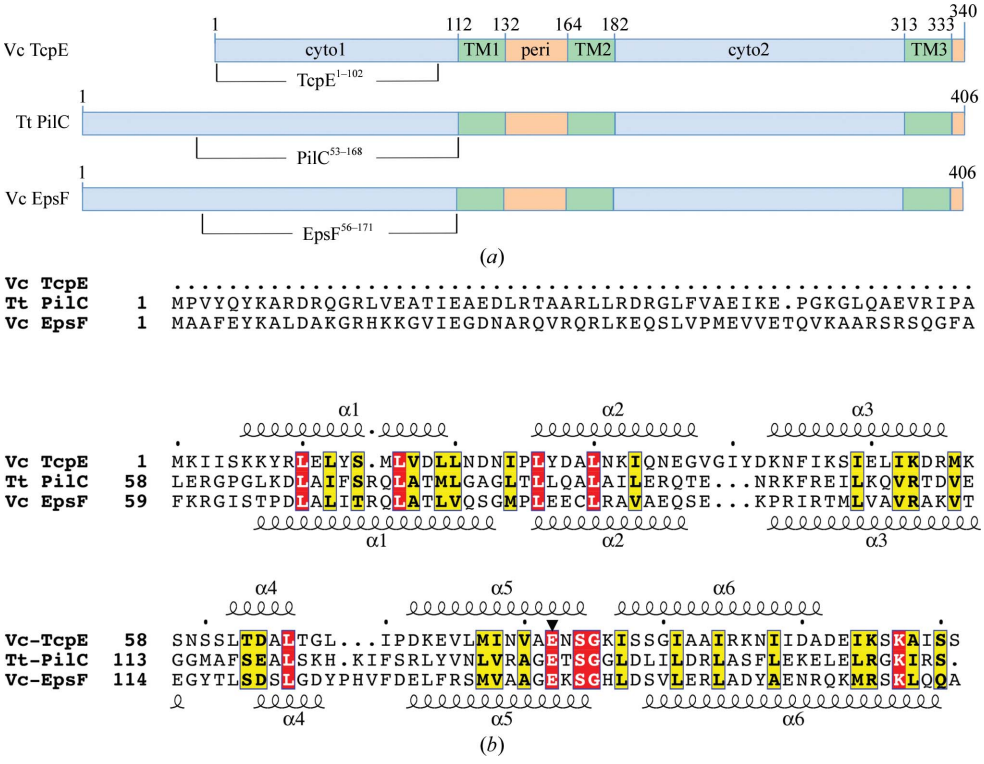

Figure 1
Putative domain and topology arrangement, sequence alignment and secondary structure of GspF family proteins. (*a*) Schematic of the predicted domain arrangement for the inner membrane core proteins (IMCPs) TcpE from the *V. cholerae* (Vc) type IVb pilus system, PilC from the *T. thermophilus* (Tt) type IVa pilus system and EpsF from the *V. cholerae* T2S system. Cyto, cytoplasmic; TM, transmembrane; peri, periplasmic. Residues that are resolved in the cyto1 crystal structures are indicated. (*b*) Amino-acid sequence alignments of the cyto1 domains. Homologous residues have a yellow background and identical residues have a red background. The conserved Glu82, which was substituted by Arg and Ala, is indicated by a solid triangle. The secondary structures are shown for *V. cholerae* TcpE above its sequence and for EpsF below its sequence. NCBI accession Nos.: Vc TcpE, AAA27563; Tt PilC, YP_144060; Vc EpsF, AAA58787.

Abendroth *et al.*, 2009) with a root-mean-square deviation (r.m.s.d.) of 1.6 Å (main-chain atoms of EpsF residues 64–161) and on PilC (PDB entry 2whn; Karuppiah *et al.*, 2010) with an r.m.s.d. of 1.5 Å (PilC residues 63–162) (Fig. 3*a*). Most of the conserved residues are hydrophobic and are oriented into the core of the six-helix bundle to stabilize this fold, with variable residues exposed on the surface of the bundle. The putative membrane-interaction surface has a net positive charge in all three N-terminal domains (Fig. 3*b*). The three proteins also share the hydrophobic groove at the bottom of the cylinder, together with the conserved residues Glu, Ser and Gly on one lip of the groove, and the intramolecular interactions formed by these residues are similar. This surface may serve as an interaction interface for one of the inner membrane accessory proteins or the assembly ATPase. To test this possibility, we substituted Glu82 in TcpE$^{1–102}$ with alanine or arginine using site-directed mutagenesis of the *tcpE* gene in

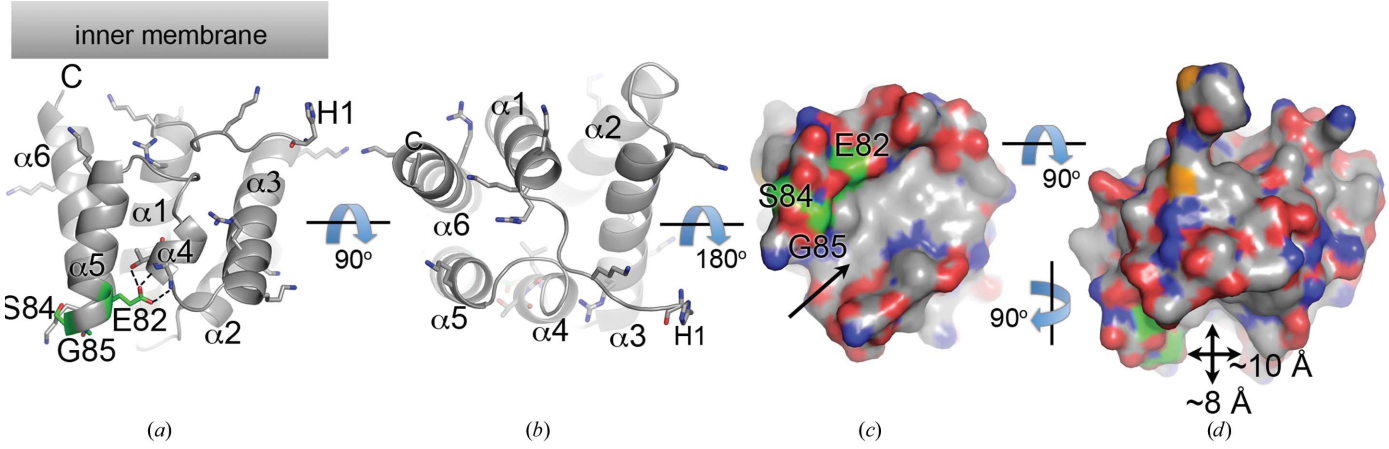

Figure 2
Crystal structure of the N-terminal cytoplasmic domain of the *V. cholerae* T4P inner membrane core protein TcpE (cyto1-TcpE) at 1.88 Å resolution. (*a*) Side view of cyto1-TcpE shown relative to the inner membrane. Lysine and arginine side chains are shown in stick representation, as are the conserved residues Glu82, Ser84 and Gly85 (behind Ser84). C atoms of the conserved residues Glu82, Ser84 and Gly85 are colored green, O atoms are red and N atoms are blue. All other atoms are gray. Dashed lines represent hydrogen bonds between Glu82 and Leu62 and Thr63. (*b*) Top view of cyto1-TcpE shown from the face of the cylinder that is predicted to interact with the inner membrane. (*c*) Surface representation of TcpE$^{1–102}$ shown from the bottom of this domain looking into the hydrophobic groove (arrow). C atoms are green for the conserved residues Glu82, Ser84 and Gly85 and gray for all other residues. O atoms are colored red and N atoms are blue for all residues. (*d*) Surface representation of TcpE$^{1–102}$ shown from a side view rotated ∼90° about the cylinder axis relative to (*a*) to show the approximate dimensions of the hydrophobic groove and the position of the conserved residues on one lip of the groove.

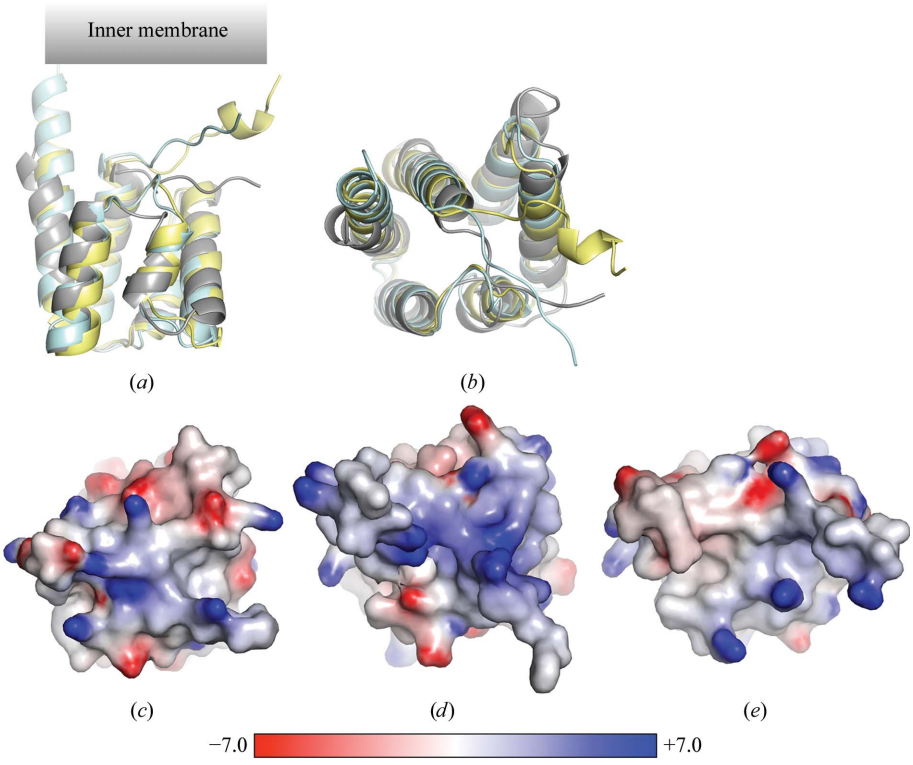

Figure 3
Comparison of the TcpE$^{1–102}$ crystal structure with the corresponding structures in EpsF and PilC. (*a*, *b*) Superposition of TcpE$^{1–102}$ (gray), EpsF$^{56–171}$ (pale cyan; PDB entry 3c1q) and PilC$^{53–168}$ (yellow; PDB entry 2whn) shown from (*a*) the side view and (*b*) the top view, looking down on the inner membrane face. (*c*–*e*) Electrostatic surface potentials for (*c*) cyto1-TcpE, (*d*) EpsF$^{56–171}$ and (*e*) PilC$^{53–168}$ shown from the top view as in (*b*). The electrostatic potential was calculated using *DelPhi* (Oron *et al.*, 2003), where red represents negative charge, blue represents positive charge and white is neutral (scale from $-7kT$ to $+7kT$).

pET-15b and expressed the TcpE$^{1–102}$ variants using the same protocol as used for wild-type cyto1-TcpE. Our intention was to alter the charge at this site to disrupt interactions with partner proteins in TcpE, but the mutation was first introduced into recombinant cyto1-TcpE to identify undesired effects on the protein fold. The cyto1-TcpE^{E82A} and cyto1-TcpE^{E82R} variants expressed well in *E. coli* RosettaBlue (DE3) but were both insoluble, suggesting that the hydrogen bonds between the Glu82 side chain and α4 (Fig. 2*a*) are critical for domain stability in TcpE and likely in other GspF family members.

3.4. Oligomerization of cyto1-TcpE

TcpE$^{1–102}$ is monomeric as a soluble protein and in its crystal form. In contrast, PilC$^{53–168}$ eluted as a dimer from a size-exclusion column (Karuppiah *et al.*, 2010) and cross-linking experiments showed that a small portion of EpsF$^{56–171}$ expressed in *V. cholerae* forms a homodimer (Abendroth *et al.*, 2009). Furthermore, both EpsF$^{56–171}$ and PilC$^{53–168}$ crystallized as dimers, which the authors proposed may be physiologically relevant. Yet, despite their structural similarity, the dimer interfaces of EpsF$^{56–171}$ and PilC$^{53–168}$ are completely different. EpsF$^{56–171}$ dimerizes about a twofold symmetry axis, with the two monomers lying in the same plane and the two α6 helices oriented ∼45° relative to each other. In the PilC$^{53–168}$ dimer one monomer is shifted downwards relative to the other such that their TM1 segments would not lie in the same plane, making this dimer an unlikely conformation for membrane interaction. While TcpE$^{1–102}$ does not form homodimers in solution or in the crystal, it is likely that this and all other IMCPs form homo-multimeric complexes that in turn interact with partner-protein multimers to form a functional inner membrane platform. More complete structures of these proteins and their complexes with partner proteins will help to clarify the quaternary structure of this filament-assembly machinery.

4. Conclusions

With our TcpE$^{1–102}$ crystal structure, atomic resolution structures are now available for IMCPs from each of the type IVa pilus, type IVb pilus and T2S systems. The close structural similarity of these domains and of the overall domain organizations of these IMCPs suggest a common role and mechanism of filament assembly for all three systems. The oligomerization state is likely to depend on the interactions of these proteins with the membrane and with other members of the assembly platform.

Portions of this research were carried out at the SSRL, a Directorate of SLAC National Accelerator Laboratory and an Office of Science User Facility operated for the US Department of Energy Office of Science by Stanford University. We thank the staff at SSRL beamline 9-2 for assistance with remote data collection and Owen Pierce for sequence analysis. Funding for this research was provided by the Canadian Institutes of Health Research (CIHR). LC was supported by a New Investigator Award from CIHR and a Scholar Award from the Michael Smith Foundation for Health Research.

References

Abendroth, J., Bagdasarian, M., Sandkvist, M. & Hol, W. G. J. (2004). *J. Mol. Biol.* **344**, 619–633.
Abendroth, J., Mitchell, D. D., Korotkov, K. V., Johnson, T. L., Kreger, A., Sandkvist, M. & Hol, W. G. J. (2009). *J. Struct. Biol.* **166**, 303–315.
Anantha, R. P., Stone, K. D. & Donnenberg, M. S. (1998). *Infect. Immun.* **66**, 122–131.
Arts, J., de Groot, A., Ball, G., Durand, E., Khattabi, M. E., Filloux, A., Tommassen, J. & Koster, M. (2007). *Microbiology*, **153**, 1582–1592.
Ayers, M., Howell, P. L. & Burrows, L. L. (2010). *Future Microbiol.* **5**, 1203–1218.

Ayers, M., Sampaleanu, L. M., Tammam, S., Koo, J., Harvey, H., Howell, P. L. & Burrows, L. L. (2009). *J. Mol. Biol.* **394**, 128–142.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Blank, T. E. & Donnenberg, M. S. (2001). *J. Bacteriol.* **183**, 4435–4450.
Bose, N. & Taylor, R. K. (2005). *J. Bacteriol.* **187**, 2225–2232.
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.* D**66**, 12–21.
Chiang, P., Habash, M. & Burrows, L. L. (2005). *J. Bacteriol.* **187**, 829–839.
Chiang, P., Sampaleanu, L. M., Ayers, M., Pahuta, M., Howell, P. L. & Burrows, L. L. (2008). *Microbiology*, **154**, 114–126.
Collins, R. F., Saleem, M. & Derrick, J. P. (2007). *J. Bacteriol.* **189**, 6389–6396.
Craig, L. & Li, J. (2008). *Curr. Opin. Struct. Biol.* **18**, 267–277.
Crowther, L. J., Anantha, R. P. & Donnenberg, M. S. (2004). *Mol. Microbiol.* **52**, 67–79.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Freitag, N. E., Seifert, H. S. & Koomey, M. (1995). *Mol. Microbiol.* **16**, 575–586.
Giltner, C. L., Nguyen, Y. & Burrows, L. L. (2012). *Microbiol. Mol. Biol. Rev.* **76**, 740–772.
González, A., Moorhead, P., McPhillips, S. E., Song, J., Sharp, K., Taylor, J. R., Adams, P. D., Sauter, N. K. & Soltis, S. M. (2008). *J. Appl. Cryst.* **41**, 176–184.
Han, X., Kennan, R. M., Parker, D., Davies, J. K. & Rood, J. I. (2007). *J. Bacteriol.* **189**, 5022–5033.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Karuppiah, V., Hassan, D., Saleem, M. & Derrick, J. P. (2010). *Proteins*, **78**, 2049–2057.
Kaufman, M. R., Seyer, J. M. & Taylor, R. K. (1991). *Genes Dev.* **5**, 1834–1846.
Kaufman, M. R., Shaw, C. E., Jones, I. D. & Taylor, R. K. (1993). *Gene*, **126**, 43–49.
Kirn, T. J., Bose, N. & Taylor, R. K. (2003). *Mol. Microbiol.* **49**, 81–92.
LaPointe, C. F. & Taylor, R. K. (2000). *J. Biol. Chem.* **275**, 1502–1510.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Lory, S. & Strom, M. S. (1997). *Gene*, **192**, 117–121.
Merz, A. J., So, M. & Sheetz, M. P. (2000). *Nature (London)*, **407**, 98–102.
Misic, A. M., Satyshur, K. A. & Forest, K. T. (2010). *J. Mol. Biol.* **400**, 1011–1021.
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.* D**67**, 355–367.
Nudleman, E., Wall, D. & Kaiser, D. (2006). *Mol. Microbiol.* **60**, 16–29.
Nunn, D. (1999). *Trends Cell Biol.* **9**, 402–408.
Oron, A., Wolfson, H., Gunasekaran, K. & Nussinov, R. (2003). *Curr. Protoc. Bioinformatics*, Unit 8.4. doi:10.1002/0471250953.bi0804s02.
Peabody, C. R., Chung, Y. J., Yen, M.-R., Vidal-Ingigliardi, D., Pugsley, A. P. & Saier, M. H. Jr (2003). *Microbiology*, **149**, 3051–3072.
Planet, P. J., Kachlany, S. C., Fine, D. H., DeSalle, R. & Figurski, D. H. (2003). *Nature Genet.* **34**, 193–198.
Pugsley, A. P. (1993). *Microbiol. Rev.* **57**, 50–108.
Sandkvist, M. (2001). *Mol. Microbiol.* **40**, 271–283.
Schmidt, S. A., Bieber, D., Ramer, S. W., Hwang, J., Wu, C.-Y. & Schoolnik, G. (2001). *J. Bacteriol.* **183**, 4848–4859.
Taylor, R. K., Miller, V. L., Furlong, D. B. & Mekalanos, J. J. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 2833–2837.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Thomas, J. D., Reeves, P. J. & Salmond, G. P. C. (1997). *Microbiology*, **143**, 713–720.
Tønjum, T., Caugant, D. A., Dunham, S. A. & Koomey, M. (1998). *Mol. Microbiol.* **29**, 111–124.
Tønjum, T., Freitag, N. E., Namork, E. & Koomey, M. (1995). *Mol. Microbiol.* **16**, 451–464.
Tripathi, S. A. & Taylor, R. K. (2007). *J. Bacteriol.* **189**, 4401–4409.
Whitchurch, C. B., Hobbs, M., Livingston, S. P., Krishnapillai, V. & Mattick, J. S. (1991). *Gene*, **101**, 33–44.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wolfgang, M., Lauer, P., Park, H.-S., Brossay, L., Hébert, J. & Koomey, M. (1998). *Mol. Microbiol.* **29**, 321–330.
Yamagata, A., Milgotina, E., Scanlon, K., Craig, L., Tainer, J. A. & Donnenberg, M. S. (2012). *J. Mol. Biol.* **419**, 110–124.
Zhang, H.-Z., Lory, S. & Donnenberg, M. S. (1994). *J. Bacteriol.* **176**, 6885–6891.