Published in final edited form as:

J Mol Biol. 2008 March 14; 377(1): 91–103.

Structure of the minor pseudopilin EpsH from the Type 2 Secretion System of *Vibrio cholerae*

Marissa E. Yanez^{a,b}, Konstantin V. Korotkov^a, Jan Abendroth^{a,\$}, and Wim G. J. Hol^{a,b,*}

^aDepartment of Biochemistry, Biomolecular Structure Center, University of Washington, Box 357742, Seattle WA 98195, USA

^bDepartment of Bioengineering, University of Washington, Box 355061, Seattle WA 98195, USA

Abstract

Many Gram-negative bacteria use the multi-protein Type II Secretion System (T2SS) to selectively translocate virulence factors from the periplasmic space into the extracellular environment. In *Vibrio cholerae* the T2SS is called the Extracellular protein secretion (Eps) system which translocates cholera toxin and several enzymes in their folded state across the outer membrane. Five proteins of the T2SS, the pseudopilins, are thought to assemble into a pseudopilus, which may control the outer membrane pore EpsD, and participate in the active export of proteins in a “piston-like” manner. We report here the 2.0 Å resolution crystal structure of an N-terminally truncated variant of EpsH, a minor pseudopilin from *Vibrio cholerae*. While EpsH maintains an N-terminal α -helix and C-terminal β -sheet consistent with the Type 4a Pilin fold, structural comparisons reveal major differences between the minor pseudopilin EpsH and the major pseudopilin GspG from *Klebsiella oxytoca*: EpsH contains a large β -sheet in the variable domain, where the GspG contains an α -helix. Most importantly, EpsH contains at its surface a hydrophobic crevice between its variable and conserved β -sheets, wherein a majority of the conserved residues within the EpsH family are clustered. In a tentative model of a T2SS pseudopilus with EpsH at its tip, the conserved crevice faces away from the helix axis. This conserved surface region may be critical for interacting with other proteins from the T2SS machinery.

Keywords

General Secretion Pathway; extra cellular protein secretion; cholera; Type 4 Pilin Biogenesis; Gsp; EpsH; PulG; Pile

Introduction

The Type II Secretion System (T2SS), also called the General Secretory Pathway (GSP), is a sophisticated multi-protein complex that selectively translocates proteins from the periplasm of Gram-negative bacteria across the outer membrane into the extracellular environment.^{1,2} Several important bacterial pathogens secrete toxins, proteases, cellulases and lipases via the T2SS.^{3–7} In *V. cholerae* the T2SS is involved in the secretion of at least five virulence factors

*Corresponding author email: wghol@u.washington.edu

\$Current address: Emerald BioStructures Inc., Bainbridge Island, WA 98110, USA

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.

including chitinase and the AB₅ heterohexamer of cholera toxin (CT), the major causative agent of the devastating disease cholera.^{8,9}

The number of proteins associated with the T2SS is species-dependent and varies from 12 to 15. Given that a large number of papers refer to T2SS proteins by a species-specific prefix, the nomenclature of the T2SS components is complex. The T2SS system of *V. cholerae*, also called the Extracellular Protein Secretion (Eps) system, consists of multiple copies of at least 11 different proteins, referred to as EpsC-EpsM.¹⁰ In the current paper, all non-*V. cholerae* T2SS proteins will be referred to by a generic “Gsp” prefix followed by a capital letter, which indicates a specific T2SS protein. For example, EpsH is homologous to GspH from a species other than *V. cholerae*.

The T2SS apparatus spans the entire cell envelope from the “protein conducting” pore in the outer membrane (the “secretin” GspD) via the pseudopilins in the periplasm (GspG-GspK) and several proteins in the inner membrane sub-complex (GspC, GspF, GspL, GpsM), to an ATPase in the cytoplasm (GspE). While several structures of Gsp proteins from the T2SS have been published in recent years,¹¹⁻¹⁶ the overall configuration of the T2SS, and the mechanisms by which unrelated proteins are exported in a folded state, remain poorly understood.^{1,2,8,17-19} Proposed mechanisms for the functioning of the T2SS include the assembly of a “pseudopilus” in the periplasm that either functions as a retractable “plug” for the outer membrane pore, or as a “piston” that actively pushes substrates through the outer membrane.²⁰⁻²³

Up to 10 proteins of the T2SS share homology with Type IV Pilus Biogenesis (T4PB) proteins,^{24,25} among these are the pseudopilins, GspG, H, I, J and K, which are related to pilins. Pilins are proteins incorporated into pili, long appendages on the surface of bacteria forming thin, strong fibers with multiple functions.^{26,27} Pilins and pseudopilins contain a pre-pilin leader sequence which is cleaved off by a pre-pilin peptidase yielding mature protein.^{28,29} The Type IV pilins (T4P's) from the T4PB system can be subdivided into Type 4a pilins (T4aP's) and Type 4b pilins. T4aP's contain a shorter pre-pilin leader sequence and exhibits within the subfamily a higher sequence homology in the hydrophobic N-terminus than the Type 4b pilin subset³⁰. Interestingly, the pseudopilins GspG-GspJ and those T4Ps that assemble into a pilus, all have a glutamate at position +5 in the mature protein.^{31,29} Given the similarities between components of the T2SS and the T4PB systems, the assembly of the T2SS pseudopilus is thought to occur in a manner analogous to the assembly of Type 4 pili^{32,33}, even though the true pili extend outside the bacterial cell wall and the pseudopili function in the periplasmic space, between the inner and outer membrane subcomplexes of the T2SS.

Several studies on T2SS-containing bacteria have demonstrated that, upon over-expression of the pseudopilin GspG^{34,22,15}, the assembly of a pilus occludes the outer membrane pore, GspD. The over-expression of the pseudopilins GspH, I, J and K, however, failed to form assembled pili.²² Furthermore, a ratio of GspG:GspH:GspI:GspJ has been estimated to be 16:1:1:4 in purified pseudopili,³⁵ implying that GspG is the predominant subunit of the T2SS pseudopilus. Consequently, GspG is referred to as the “major” pseudopilin, while GspH, I, J and K are called “minor” pseudopilins.³⁵ The crystal structure of GspG from *Klebsiella oxytoca*,¹⁵ which shares 80% sequence identity with EpsG from *V. cholerae*, revealed that this major pseudopilin adopts the Type 4a Pilin fold,³⁰ consisting of an N-terminal α -helix, a variable region and a “conserved β -sheet”, consisting of four anti-parallel β strands.

The function and localization of minor pseudopilins are still to be fully unraveled. It has been suggested that some minor pseudopilins may assemble either into the base or the tip of pili, or both, and function as initiators or regulators of pilus biogenesis and dynamics, and/or as adaptors between various pseudopilin components and other members of the T2SS.^{22,36-39}

Several studies have demonstrated that all four minor pseudopilins are essential for proper functioning of the T2SS.^{40,22,39} However, the results of mutagenesis studies on GspH homologues have varied slightly among species, as deletion of the *gspH* gene from *K. oxytoca* decreased secretion to ~20% of normal,³⁹ while deletion of the *gspH* gene from *Erwinia chrysanthemi* completely abolished secretion.⁴⁰

We report here the crystal structure of EpsH from *V. cholerae* refined to 2.0 Å resolution, which represents the first atomic structure of a minor pseudopilin from a T2SS. The minor pseudopilin EpsH differs significantly from the major pseudopilin GspG from *K. oxytoca*. It contains the largest β -sheet in the variable pilin domain seen among T4aP-like structures to date. Despite low overall sequence identity among the EpsH/GspH family members, we were able to identify a cluster of highly conserved residues that form a hydrophobic cleft between the variable and conserved β -sheets. This conserved surface area may be the site where EpsH/GspH pseudopilins interact with one or more other T2SS proteins. In a tentative model of a T2SS pseudopilus with EpsH at its tip, the conserved crevice faces away from the helix axis.

Results

The Structure of EpsH

An N-terminally truncated soluble variant of EpsH starting at residue +30 of the mature sequence with a C-terminal hexa-histidine tag yielded well diffracting crystals. The resultant crystal structure of EpsH³⁰⁻¹⁸⁸ could be determined using Se-Met single-wavelength anomalous dispersion (SAD) phasing and refined to 2.0 Å resolution with $R_{\text{work}} = 16.9\%$, $R_{\text{free}} = 21.2\%$ and good geometry (Table 1). The structure determination was challenging probably due to the fact that there was only one well-ordered methionine per 159 residues. Two monomers of EpsH³⁰⁻¹⁸⁸ were present in the asymmetric unit (ASU) (Figure 1). Residues Asp30-Asp122 and Glu136-Asp186 of monomer A, and Asp30-Leu104 and Glu136-Asp186 of monomer B could eventually be built into the electron density.

Each monomer of EpsH³⁰⁻¹⁸⁸ contains a hydrophobic N-terminal α -helix (Asp30-Leu54), which is characteristic of all Type 4 pilins, followed by nine β -strands forming two β -sheets: one 5-stranded anti-parallel β -sheet (β -sheet I) composed of strands β 1, β 2, β 3, β 4 and β 6, and one 4-stranded anti-parallel sheet (β -sheet II) composed of strands β 5, β 7, β 8 and β 9 (Figure 1). A long extended loop, connecting β 5 from β -sheet II to β 6 of β -sheet I, is comprised of residues Thr92-Gln139, of which 13 residues in monomer A and 32 residues in monomer B are disordered. A loop following strand β 6 connects this final strand from β -sheet I to β 7 from β -sheet II (Figures 1, 2).

Intermolecular Interactions

A total of ~1370 Å², or 17% of the total solvent accessible surface area of monomer A of EpsH³⁰⁻¹⁸⁸ in the crystal, is buried by intermolecular contacts with three adjacent EpsH monomers. The first interface is formed between the anti-parallel N-terminal helices of two monomers A and B within the ASU (interface A-B Figure 1(b)) burying ~820 Å² solvent accessible surface with a gap volume index (GVI)⁴¹ of 4.94 Å. Here, 13 residues in monomer A contact the same residues in monomer B and engage primarily in hydrophobic interactions with only one H-bond between Leu52 in monomer A and Gln37 in monomer B. The second interface, between monomers A and B' (interface A-B' in Figure 1(b)), buries a surface area of ~770 Å², slightly smaller than the A-B interface, with a GVI of 2.03 Å. Here, 16 residues in monomer A contact 8 residues in monomer B' forming a large number of hydrophobic interactions and two H-bonds: one between Glu162 in monomer A and Glu162 in monomer B', and one between Gln166 in monomer A and Asp164 in monomer B'. The third and largest intermolecular interface occurs between monomers A and B'' (interface A-B'' in Figure 1(b)),

which buries a total of $\sim 1050 \text{ \AA}^2$ with a GVI of $2.51 \text{ \AA}$. Here, 17 residues in monomer A and the same 17 residues in monomer B” form hydrophobic contacts and two H-bonds, one between Ala75 in monomer A and Thr150 in monomer B”, and vice versa. The last two interactions have GVI values in the range expected for physiological dimers^{41,42} and the significance of these crystal contacts will be discussed below.

Sequence homology

The nine closest sequence homologues of *V. cholerae* EpsH (VcEpsH), as identified by the BLAST program⁴³, are from *V. vulnificus* (50% sequence identity with EpsH), *V. parahaemolyticus* (50%), Enterotoxigenic *E. coli* (35%), *Shigella dysenteriae* (28%), *Aeromonas hydrophila* (28%), *K. oxytoca* (25%), *Erwinia carotovora* (24%) and *Erwinia chrysanthemi* (19%) (Figure 2). Hence, the pairwise protein sequence identity within these EpsH/GspH homologues varies between 19-50%. In contrast, the major pseudopilin EpsG/GspG family has a much higher pairwise sequence homology, which varies between 60-80% for the same species used for the EpsH/GspH comparison. Interestingly, the sequence homology of EpsH/GspH is consistent with what is typically seen in Type 4 pilins,³⁰ showing a conserved hydrophobic N-terminal α -helix followed by a highly variable region (both with respect to length and secondary structure), and a moderate degree of homology in the C-terminal region.

The multiple sequence alignment of the EpsH/GspH family shows that out of the 159 residues in our EpsH³⁰⁻¹⁸⁸ structure, seven residues (Phe39, Gly58, Pro139, Pro151, Phe152, Leu154, Gly174) are invariant (labeled red in Figure 2), and an additional four residues (Gly54, Gly147, Glu148, Thr150) are highly homologous, defined as differing only in one or two of the nine EpsH/GspH homologues (pink in Figure 2). The EpsH/GspH alignment shows an additional 18 residues that are conserved as a hydrophobic residue in all GspH variants (red letters in Figure 2).

Interestingly, a majority of the conserved residues in the EpsH/GspH family are clustered around a specific region of the EpsH structure at the point of convergence of the two β -sheets and the N-terminal helix (Figure 3):

- i. A conserved stretch of sixteen residues (Pro139-Leu154) lies along the C-terminal strand $\beta 6$ of β -sheet I, the loop that connects $\beta 6$ to $\beta 7$ of β -sheet II, and part of strand $\beta 7$ (Figure 2-red box). Of these sixteen residues, ten are highly homologous in the EpsH/GspH family. Four out of the ten are completely conserved (Pro139, Pro151, Phe152 and Leu154), three show homologous substitutions (Gly147, Glu148, Thr150), and three conserve a hydrophobic character in all GspH variants (Leu141, Val143, Val149);
- ii. Three additional residues, provided by the N-terminal α -helix, are spatially located nearby. One of these residues is completely conserved (Phe39) and two maintain a hydrophobic character (Leu43, Leu46) in the EpsH/GspH family. These latter three residues form the “base” of this region between β -sheet I and β -sheet II;
- iii. One completely conserved residue (Gly58) and three hydrophobic residues (Phe57, Val59, Ile61) from strand $\beta 1$ of β -sheet I form the “left” (as viewed in Figure 3C) edge of this region (Figure 3(c)).

In summary, fifteen of the seventeen conserved EpsH/GspH residues clustered within this region are hydrophobic. This cluster of conserved residues forms a solvent-exposed hydrophobic groove (Figure 3).

Electrostatics

The electrostatic potential of the surface of EpsH³⁰⁻¹⁸⁸, as calculated by APBS⁴⁴ within Pymol⁴⁵, revealed that the “back side” of EpsH (not shown) has a rather diffuse charge distribution. The “front side”, however, has a significant region of negative charge located primarily on the side of EpsH where the conserved β -sheet II is located (Figure 3(b)). Here, nine residues within the loops between strands β 5, β 7, β 8 and β 9 of β -sheet II (Asp111, Asp112, Glu148, Glu162, Asp164, Glu165, Glu173, Glu184, Asp185) and two residues along β 5 (Asp100, Glu102) form a large area of negative charge. On the opposite side of the molecule, a small patch of positive charge is present in the variable β -sheet I, formed by three residues, Arg60, Lys81 and Lys86. A comparison of the electrostatic and sequence homology surface representations of EpsH³⁰⁻¹⁸⁸ (Figures 3(a) and 3(b)) shows, however, that these two regions with positive and negative charge have a low degree of sequence homology within the T2SS family.

Structural Comparison with a Major Pseudopilin

Substantial structural differences appear to exist between the minor pseudopilin EpsH³⁰⁻¹⁸⁸ and the major pseudopilin GspG from *K. oxytoca* (PDB code 1T92)¹⁵ (Figure 4). The differences are so extensive that the DALI server⁴⁶ failed to recognize the *K. oxytoca* GspG structure as an EpsH homolog. Both EpsH and *K. oxytoca* GspG have an N-terminal α -helix and a C-terminal β -sheet, with the only β -sheet in *K. oxytoca* GspG corresponding to β -sheet II in EpsH. However, the strands of β -sheet II in EpsH are longer than the strands of the corresponding β -sheet in GspG.

The variable regions are entirely different in the two proteins. The variable domain of *K. oxytoca* GspG (residues 55-93) consists of a long, irregular loop and one α -helix of one turn (residues 65-67), while the variable domain of EpsH consists of the extensive β -sheet I. A global superposition using the entire structure of each protein with the program SSM was only able to superimpose 38 residues with an r.m.s.d. of 2.6 Å and a sequence identity of 2.6%. When comparing only the C-terminal β -sheets in these two proteins, 26 residues were superimposed with an r.m.s.d. of 1.5 Å and a sequence identity of 8%. The latter superposition revealed a large difference in the angle between the N-terminal α -helix and the conserved β -sheet for EpsH vs. *K. oxytoca* GspG. Remarkably, there appears to be only a superficial similarity in the folds of the T2SS pseudopilins *V. cholerae* EpsH and *K. oxytoca* GspG.

Structural Comparison with Type 4a Pilins

At the time of this study, the structures of three T4aP's were deposited in the protein data bank: PilA^{PAK} from *Pseudomonas aeruginosa* strain PAK (PDB code 1DZO)⁴⁷, PilA^{K1224} from *P. aeruginosa* strain K122-4 (PDB code 1HPW)⁴⁸, and PilE from *Neisseria gonorrhoeae* (PDB code 2HI2)⁴⁹ (Figure 4(a) and 4(b)). All three structures are from major pilins of their respective T4PB systems. Of these, PilA^{PAK} is the T4aP with the closest structural homology to EpsH. Surprisingly, structural comparisons revealed a higher structural homology between the minor pseudopilin EpsH and the major Type 4a Pilin PilA^{PAK} than between the major and minor pseudopilins GspG and EpsH. Like EpsH, PilA^{PAK} exhibits a β -sheet in the variable domain, though, with only 14 residues, it is significantly smaller than the variable β -sheet I in EpsH, which is composed of 21 residues. A global alignment with SSM was able to superimpose 67 residues out of 143 EpsH residues onto PilA^{PAK} with an r.m.s.d. of 2.5 Å and a sequence identity of 15% (Figure 4c). The resultant structure-based sequence alignment shows that 10 out of the aligned 67 residues are identical in EpsH and PilA^{PAK} (Figure 4(c)). Those residues include one proline (Pro139 in EpsH), which is conserved in all EpsH/GspH homologues and which is part of the conserved hydrophobic surface patch of EpsH (Figure 3(c)). Two glycines, Gly175 in EpsH and Gly125 in PilA^{PAK}, are conserved in all EpsH/GspH and PilA homologues. The region of the highest sequence homology between EpsH and PilA^{PAK} is

indicated by the pink box in Figure 4(c). Interestingly, this area corresponds to the same region between the two β -sheets of EpsH that is most highly conserved in the EpsH/GspH family (pink box in Figure 2; Figure 3). The two black arrows shown in Figure 4(c) point to the two prolines (Pro89 and Pro91) which form a two cis-proline containing loop in PilA^{PAK}. The two corresponding prolines in EpsH are, however, in a *trans* rather than a *cis* conformation.

All five Type IVa Pilin-like structures that have been solved to date, i.e. the three major pilins from the T4PB system, and a minor and major pseudopilin from the T2SS, are shown in Figure 4. The same structural features that are shared between EpsH and *K. oxytoca* GspG (i.e. a hydrophobic extended N-term α -helix and the C-terminal β -sheet) are conserved among all five Type 4a-like Pilins with known structures. The conserved β -sheet of the two pseudopilins share the same anti-parallel topology as in the three Type IVa Pilins with known structure (Figure 4(b)). A superposition of all five pilin structures showed that only 26-32 residues within the structurally conserved β -sheet of all five proteins can be superimposed with pairwise r.m.s. deviations ranging from 1.5 - 2.0 Å. The superposition revealed a major difference in the angle by which the helix of each Type 4a-like pilin is oriented relative to the conserved β -sheet (Figure 4(a)). In comparison to the four other Type 4a-like pilin structures, EpsH has the most extensive variable pilin domain seen to date in the Type 4a Pilin family.

The N-terminal residues of EpsH

Truncating the first 24-26 N-terminal residues is an approach that has been used to solve several Type 4a pilin structures, such as PilA^{PAK} 47,48; a major pseudopilin PulG¹⁵; and two Type 4b pilin structures, TcpA³⁰ and BfpA⁵⁰. Full length structures of Type 4a-like Pilins have shown that their N-termini form long α -helices^{51,30,49}, which are very likely packed near the center of the Type 4 Pilus⁴⁹ and of the T2SS pseudopilus¹⁵. It is generally believed that a long N-terminal α -helix is a feature of the entire Type 4a pilin and pseudopilin family^{52,53}. Given that PilA^{PAK} is the closest structural homolog of *V. cholerae* EpsH regarding the globular domains of these two proteins (Fig. 4b), it is of interest to compare their N-termini. They appear to be very similar in sequence (Fig. 4d). It is of particular interest that the canonical Glu5 is maintained and that the starting five residues are identical between PilA^{PAK} and EpsH. Also the nature of the next 16 residues - almost all having small and aliphatic side chains - is well maintained in these two proteins. It is therefore likely that they adopt quite similar α -helical structures. The observed N-terminal helix α -1 in our current structure of truncated EpsH can be extended without any problems by 29 helical residues towards the N-terminus of the mature protein. This suggests that the packing of EpsH with partner pseudopilins may resemble the packing of other pilins and pseudopilins in Type 4 Pili and T2SS pseudopili.

Discussion

Oligomeric state of EpsH

Since yeast two-hybrid studies indicated interactions between N-terminally truncated GspH proteins from *Erwinia chrysanthemi*⁵⁴, and also since pilins make extensive intersubunit contacts within an assembled pilus,^{15,49} it is of interest to investigate the interactions between adjacent monomers of *V. cholerae* EpsH in our crystals (Figure 1b).

For the A-B interface, a GVI of 4.94 Å indicates an extremely loose contact between monomers A and B of EpsH³⁰⁻¹⁸⁸ (Ref. 41). Moreover, the helices of monomers A and B of EpsH³⁰⁻¹⁸⁸ are oriented with respect to one another in an anti-parallel direction, which is the opposite of what was observed in the cryo-electron microscopy structure of the PilE pilus from *N. gonorrhoeae*.⁴⁹ Furthermore, of the four residues that are most buried in the A-B interface (Gln41, Leu44, Glu48, Leu52), none are completely conserved within the 10 closest EpsH

homologues (Figure 2). Therefore, the A-B dimer is probably a crystal-packing contact, not occurring in an assembled T2SS.

For the second interface, a GVI of 2.03 Å indicates a much tighter fit between monomers A and B' than between monomers A and B, although the buried surface area of ~770 Å² in the A-B' interface is relatively small. However, there is very low sequence homology among contact residues in the A-B' interface (Fig.2). Also, the observed interaction between the carboxylates of Glu 162 in monomer A and Glu162' in monomer B requires the low pH of 4.5 of the crystallization medium. Therefore, the A-B' contacts observed in the crystals are not likely to be physiological relevant.

The largest interface in the crystals occurs between monomers A and B'' and has a GVI of 2.51 Å, well within the range seen for physiological homodimers.⁴² Additionally, there is a significantly degree of sequence homology among residues in the A-B'' interface (pink triangles and stars in Figure 2). Yet, the contact residues that are most buried in the A-B'' interface are not well conserved. In particular, Ala75 in EpsH varies between Ala, Arg, His, and Glu among the EpsH homologues (Figure 2). Therefore, the A-B'' dimer is not likely a representation of a physiological EpsH dimer. However, we will see in the next section that the surface involved in this crystal contact is nevertheless likely of great functional interest.

Conserved Hydrophobic Crevice

An extensive hydrophobic conserved crevice, comprised mainly of 17 conserved residues, is located between the N-terminal helix, β-sheet I and β-sheet II of EpsH³⁰⁻¹⁸⁸ (Figure 3). The same region also has the highest sequence and structural homology between EpsH and PilA^{PAK} (Figure 4c). Interestingly, several of these conserved residues are involved in the A-B'' crystal contact together with non-conserved residues. This crystal contact may reflect a general propensity of the conserved patch to interact with another protein interface.

Promiscuous protein binding patches have been observed in other cases, such as in the ARF family of proteins⁵⁵ and among the ankyrin-repeat proteins⁵⁶. Taking all evidence together, the hydrophobic patch conserved within the EpsH/GspH family (Figure 3) is likely to be involved in contacting other proteins, and hence of functional significance. This area might engage in transient interactions and function e.g. as a chaperone during assembly of the T2SS, in particular of its pseudopilus. It cannot be excluded that this hydrophobic surface area of EpsH may interact transiently with one or more secreted proteins given the challenge of unraveling the interactions between T2SS components and secreted proteins¹⁷⁻¹⁹.

Alternatively, this EpsH surface area could play a role in constitutive interactions with partners in the T2SS machinery which would be in agreement with: (i) the observations of Nunn and Lory³⁵ that XcpU, the GspH homolog of *P. aeruginosa*, is present in purified pseudopili; (ii) the Ni-affinity chromatography results of Hu *et al*³⁸ indicating that in *Xanthomonas campestris* GspH interacts with GspG; (iii) the two-hybrid studies of Douet *et al*⁵⁴ which indicate that the GspH of *Erwinia chrysanthemi* interacts with itself; and, (iv) the biochemical studies of Kuo *et al*³⁷ which indicate that in *X. campestris* GspH interacts with GspI and GspJ.

Possible Position of EpsH in the T2SS Pseudopilus

With the structure of the EpsH in hand, it is tempting to speculate where in a T2SS pseudopilus the minor pseudopilin EpsH might fit. There are two the Type 4-like Pilin models described in the literature, both obtained from cryo-electron microscopy studies. One is for GspG from *K. oxytoca*, with a resolution of 25 Å, and the second for PilE from *N. gonorrhoeae*, with a resolution of 12 Å¹³. Given the difference in resolution, the PilE pilin model (PDB code 2HIL) was used for our analysis. There are some limitations for generating a reasonably accurate model of a T2SS pseudopilus including: (i) the distant relationship between PilE and major T2SS pseudopilins like *E. oxytoca* GspG with only 21% amino acid sequence identity; and (ii)

the relatively low resolution of the PilE pilus model. On the basis of an EpsG pseudopilus model, obtained as described in Materials and Methods, it appears that EpsH, when placing its helix at the “canonical” position near the center of the fiber, would not likely be able to fit at the base of the pseudopilus, due to steric hindrance of the large variable β -sheet I of EpsH and the N-terminal helix of a neighboring pseudopilin monomer. An EpsH location at the tip of the pseudopilus seems possible since in that case no serious clashes occur between EpsG subunits and EpsH. In this tentative model, the conserved hydrophobic patch on EpsH is completely surface exposed at the tip of the pseudopilus pointing away from the helix axis (Figure 5). This would enable this patch to interact with other proteins while being incorporated into the pseudopilus. Clearly, further biochemical and structural studies will have to be carried out to reveal which proteins, or possibly even small molecule partner(s), interact with EpsH, and how, in the fully assembled and functioning T2SS.

Materials and Methods

Expression and purification of EpsH³⁰⁻¹⁸⁸

The 477 bp ORF encoding the pseudopilin EpsH was PCR-amplified from *V. cholerae* genomic DNA as the template using high-fidelity KOD HiFi DNA polymerase (Novagen), a sense primer 5'-GAGATATACCATGGATGAAGCCAAATCAGTGCTCAG-3' containing an *NcoI* site (shown in bold) and an antisense primer 5'-CCCTCGAGCTCTTCATCACTTTCTCCCG-3' containing a *XhoI* site (shown in bold). The PCR-amplified fragment was digested with *NcoI* and *XhoI* and cloned into pET-21d(+) vector (Novagen) previously digested by the same restriction enzymes. The final plasmid construct encodes EpsH³⁰⁻¹⁸⁸ with an additional 8 amino acids from the vector, containing a non-cleavable hexahistidine tag.

The recombinant plasmid was transformed into *E. coli* BL21(DE3) cells and transformants were selected on LB agar plates containing 100 $\mu\text{g ml}^{-1}$ ampicillin. A single colony was picked and 50 ml primary culture was grown overnight at 310 K. Selenomethionine-labeled EpsH³⁰⁻¹⁸⁸ was obtained by first inoculating 2 ml of the overnight culture into 1 L of LB containing 50 mg/ml of carbenicillin. Cells were grown at 305 K for 3 hours to $A_{600}=0.6$ and pelleted by centrifugation at 5000 g for 15 minutes. Pellets were resuspended into M9 minimal medium containing 50 mg/ml of carbenicillin supplemented with selenomethionine and other amino acids suppressing methionine biosynthesis. Cells were induced with 0.5 mM isopropyl β -D-thiogalactopyranoside (IPTG).

Cells were harvested after 4 hours at 300K by centrifugation and resuspended in lysis buffer containing 20 mM Tris-HCl (pH 7.8), 250mM NaCl and a Complete EDTA-free Protease Inhibitor Cocktail (Roche). Cells were lysed with French Press and centrifuged at 20,000 g for 40 min. For immobilized metal affinity chromatography, Ni-NTA resin (Qiagen) was equilibrated with wash buffer containing 20 mM imidazole, 20 mM Tris-HCl (pH 7.8), 250 mM sodium chloride. The cleared lysate was loaded onto a Ni-NTA column and washed with the same buffer used for equilibration. The target protein was then eluted with 250 mM imidazole, 20 mM Tris-HCl (pH 7.8), 250 mM sodium chloride. To remove imidazole, the protein was dialysed for 24 h against 20 mM Tris-HCl (pH 7.8), 250 mM sodium chloride. The protein was concentrated using a 10 kDa molecular-weight cutoff Centricon (Amicon). For further purification, 1 ml samples of 10 mg ml^{-1} protein solution were loaded onto a Superdex 75 HR10/30 size-exclusion column (GE Healthcare) equilibrated with 20 mM Tris-HCl (pH 7.8), 250 mM NaCl, 1 mM EDTA.

Protein crystallization and X-ray diffraction data collection

Purified EpsH³⁰⁻¹⁸⁸ protein was concentrated to 10-12 mg/ml for protein crystallization. Initial crystallization conditions were searched for using the Crystal Screen I (Hampton Research) by mixing equal amounts of protein and reservoir (1.0 μ l each) using the sitting-drop variant of the vapor-diffusion method at room temperature (293K). Crystals appeared after 1 week in 0.2 M ammonium acetate, 0.1 M sodium acetate pH 4.6, 30% w/v PEG 4000. Conditions were optimized using the Additive Screen (Hampton Research) by adding 1 μ l of additive to 1 μ l of protein and 1 μ l of reservoir. Larger crystals appeared overnight using 100mM glycine as additive. Crystals were optimized at a protein concentration of 12 mg/ml in a final condition of 0.1 M sodium acetate pH 4.5-4.6, 20-25% PEG 4000, 100 mM glycine. Crystals were transferred to a cryoprotectant solution containing 1 mM EDTA, 0.25 M NaCl, 0.1 M Tris-HCl pH 7.8, 0.1 M sodium acetate pH 4.6, 0.2 M ammonium acetate, 25% PEG 4000, 10% glycerol, and flash-frozen in liquid nitrogen.

X-ray diffraction data were collected at SSRL BL9-2 at a temperature of 100 K maintained by a cold nitrogen-gas stream, using a MAR345 imaging-plate detector and a wavelength of 0.97903 Å. An EXAFS scan was performed to optimize the wavelength for the anomalous signal. Data were integrated and scaled using HKL2000 using DENZO and SCALEPACK⁵⁷ to a resolution of 2.0 Å.

Structure solution and refinement

The structure of EpsH³⁰⁻¹⁸⁸ was solved by Se-SAD using *SOLVE*⁵⁸, which located one of two methionine residues in each of the 159 residue monomers per asymmetric unit (ASU). *RESOLVE*⁵⁸ was used for initial density modification and automatic model building. In the resultant relatively poor initial electron density only 2 helices could be placed in the ASU, which were then used to derive an initial NCS operator. Twofold averaging in *RESOLVE* greatly improved the quality of the electron density map. This averaged map was used for automatic model building in *Arp/Warp*⁵⁹, which placed automatically 222 out of the 334 total amino acids in the ASU. The protein model was improved by visual inspection and manual model building using the graphics program *Coot*⁶⁰. The program *REFMAC5*⁶¹ was used for TLS refinement with 8 TLS groups, which were determined by the TLS motion determination (TLSMD) server⁶². Model building with *Coot* alternated with refinement by the program *REFMAC5*. The progress of refinement was monitored by calculation of R_{free} using 5% of the independent reflections. Residues Glu123-Gln135 in monomer A, Gly104-Gln135 in monomer B, and Pro185-Asp186 in monomers A and B were disordered and could not be included in the final structure. The stereochemical quality of the model was verified using *PROCHECK*⁶³. Final structure determination and refinement statistics are listed in Table 1.

Sequence and Structural analysis of EpsH

A multiple sequence alignment was made using the *Multalin* server⁶⁴ and the *ESPript* server was used to render the alignment.⁶⁵ The program *DSSP*⁶⁶ was used to assign the secondary-structure. The *Protein-Protein Interaction Server*⁴² was utilized to calculate buried surface areas (Δ ASA) and to identify hydrophobic contacts and hydrogen bonds (between adjacent molecules in the crystal lattice. Polar atoms within a distance range of 2.5-3.5 Å with proper hydrogen-bond geometry were considered to form hydrogen bonds. The gap volume index (GVI) is defined as: $\text{GVI} (\text{\AA}) = \text{gap volume between molecules} (\text{\AA}^3) / \text{interface ASA} (\text{\AA}^2)$ (per complex), and was calculated by the *Protein-Protein Interaction Server*. The structure superpositions were carried out using the *Secondary Structure (SSM) Superposition Method*⁶⁷ from the *CCP4* suite. Sequence conservation plotted onto the surface of EpsH³⁰⁻¹⁸⁸ structure was determined using the *CONSURF*⁶⁸ server. All figures were generated using *PYMO*L.⁴⁵

EpsH and a tentative model of a T2SS pseudopilus

Previous electron microscopy studies have indicated that Type 4a Pili from the T2SS and the T4PB system are similar in size.²⁵ A global superposition of GspG (PDB code 1T92)¹⁵, the major pseudopilin from the *K. oxytoca* T2SS, onto one monomer of PilE (PDB code 2HIL)⁴⁹, the major Type 4a pilin from the *N. gonorrhoeae* T4PB system, aligned 55 Ca atom pairs (out of a total of 99 residues for *K. oxytoca* GspG) with a r.m.s.d. of 2.0 Å and 12.7% sequence identity. All secondary structure elements of *K. oxytoca* GspG (α 1, α 2, β 1, β 2, β 3, β 4) superimpose onto equivalent secondary structure elements in PilE. Lacking from *K. oxytoca* GspG, is the so-called C-terminal “D-region” of PilE consisting of the final two C-terminal strands β 5 and β 6 in PilE.

From the crystal structure of *K. oxytoca* GspG¹⁵ and the cryo electron microscopy model of the PilE pilus from *N. gonorrhoeae*⁴⁹, we constructed a model of the assembled EpsG pseudopilus by, first, creating an EpsG homology model using the Swiss-model server (www.expasy.org/spdbv/)⁶⁹ and the *K. oxytoca* GspG structure which shares 80% sequence identity, and second, superimposing 21 residues (Ala31-Leu52) from the N-terminal α 1 of the EpsG model onto equivalent residues (Ala31-Leu52) from the N-terminal α 1 of each monomer (chains A-R) within the assembled PilE pilus (PDB code 2HIL).

To investigate possible interactions of EpsH with the pseudopilus, 21 residues (Glu31-Leu52) from the N-terminal α 1 of EpsH were then superimposed onto equivalent residues (Ala31-Leu52) of *K. oxytoca* GspG subunits at the bottom (Chain I), middle (Chain D), and top (Chain R) of the pilus model. The superposition of EpsH onto the bottom (Chain I) and middle (Chain D) subunits showed significant steric clashes between the large “variable” β -sheet I of EpsH and the N-terminal helix of adjacent subunits (Chain F and Chain A, respectively). When EpsH is superimposed onto the top subunit (Chain R) of the EpsG pseudopilus model, the bottom of the EpsH monomer has only a limited number of close contacts with the top of the EpsG subunit directly below EpsH at the position of Chain N. The EpsH-EpsG interface buries 746 Å² solvent accessible surface with a gap volume index of 2.5 Å at the tip of our preliminary pseudopilus model.

Acknowledgements

We acknowledge Stewart Turley for help with data collection, Brian Krumm for helpful discussions about crystallization, and Francis Athappilly for maintenance of our computing network. We thank Maria Sandkvist for advice, and the staff of beamline 9.2 at the SSRL, Stanford, for support during data collection. This research was supported by grant NIH grant AI34501 to W.G.J.H. from the National Institute of Allergy and Infectious Diseases, and by the Howard Hughes Medical Institute (HHMI).

Protein Data Bank accession codes

The atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank and are available under accession code 2QV8.

References

1. Johnson TL, Abendroth J, Hol WG, Sandkvist M. Type II secretion: from structure to function. *FEMS Microbiol Lett* 2006;255:175–186. [PubMed: 16448494]
2. Cianciotto NP. Type II secretion: a protein secretion system for all seasons. *Trends Microbiol* 2005;13:581–588. [PubMed: 16216510]
3. Gerritse G, Ure R, Bizoullier F, Quax WJ. The phenotype enhancement method identifies the Xcp outer membrane secretion machinery from *Pseudomonas alcaligenes* as a bottleneck for lipase production. *J Biotechnol* 1998;64:23–38. [PubMed: 9823657]
4. DeShazer D, Brett PJ, Burtnick MN, Woods DE. Molecular characterization of genetic loci required for secretion of exoproducts in *Burkholderia pseudomallei*. *J Bacteriol* 1999;181:4661–4664. [PubMed: 10419967]

5. Reeves PJ, Whitcombe D, Wharam S, Gibson M, Allison G, Bunce N, Barallon R, Douglas P, Mulholland V, Stevens S, et al. Molecular cloning and characterization of 13 out genes from *Erwinia carotovora* subspecies *carotovora*: genes encoding members of a general secretion pathway (GSP) widespread in gram-negative bacteria. *Mol Microbiol* 1993;8:443–456. [PubMed: 8326859]
6. Rossier O, Starkenburg SR, Cianciotto NP. *Legionella pneumophila* type II protein secretion promotes virulence in the A/J mouse model of Legionnaires' disease pneumonia. *Infect Immun* 2004;72:310–321. [PubMed: 14688110]
7. Davis BM, Lawson EH, Sandkvist M, Ali A, Sozhamannan S, Waldor MK. Convergence of the secretory pathways for cholera toxin and the filamentous phage, CTXphi. *Science* 2000;288:333–335. [PubMed: 10764646]
8. Sandkvist M. Type II secretion and pathogenesis. *Infect Immun* 2001;69:3523–3535. [PubMed: 11349009]
9. Sandkvist M, Michel LO, Hough LP, Morales VM, Bagdasarian M, Koomey M, DiRita VJ, Bagdasarian M. General secretion pathway (eps) genes required for toxin secretion and outer membrane biogenesis in *Vibrio cholerae*. *J Bacteriol* 1997;179:6994–7003. [PubMed: 9371445]
10. Sandkvist M, Bagdasarian M, Howard SP. Characterization of the multimeric Eps complex required for cholera toxin secretion. *Int J Med Microbiol* 2000;290:345–350. [PubMed: 11111909]
11. Robien MA, Krumm BE, Sandkvist M, Hol WG. Crystal structure of the extracellular protein secretion NTPase EpsE of *Vibrio cholerae*. *J Mol Biol* 2003;333:657–674. [PubMed: 14556751]
12. Abendroth J, Rice AE, McLuskey K, Bagdasarian M, Hol WG. The crystal structure of the periplasmic domain of the type II secretion system protein EpsM from *Vibrio cholerae*: the simplest version of the ferredoxin fold. *J Mol Biol* 2004;338:585–596. [PubMed: 15081815]
13. Abendroth J, Murphy P, Sandkvist M, Bagdasarian M, Hol WG. The X-ray structure of the type II secretion system complex formed by the N-terminal domain of EpsE and the cytoplasmic domain of EpsL of *Vibrio cholerae*. *J Mol Biol* 2005;348:845–855. [PubMed: 15843017]
14. Abendroth J, Bagdasarian M, Sandkvist M, Hol WG. The structure of the cytoplasmic domain of EpsL, an inner membrane component of the type II secretion system of *Vibrio cholerae*: an unusual member of the actin-like ATPase superfamily. *J Mol Biol* 2004;344:619–633. [PubMed: 15533433]
15. Kohler R, Schafer K, Muller S, Vignon G, Diederichs K, Philippsen A, Ringler P, Pugsley AP, Engel A, Welte W. Structure and assembly of the pseudopilin PulG. *Mol Microbiol* 2004;54:647–664. [PubMed: 15491357]
16. Korotkov KV, Krumm B, Bagdasarian M, Hol WG. Structural and functional studies of EpsC, a crucial component of the type 2 secretion system from *Vibrio cholerae*. *J Mol Biol* 2006;363:311–321. [PubMed: 16978643]
17. Bouley J, Condemine G, Shevchik VE. The PDZ domain of OutC and the N-terminal region of OutD determine the secretion specificity of the type II out pathway of *Erwinia chrysanthemi*. *J Mol Biol* 2001;308:205–219. [PubMed: 11327762]
18. Connell TD, Metzger DJ, Wang M, Jobling MG, Holmes RK. Initial studies of the structural signal for extracellular transport of cholera toxin and other proteins recognized by *Vibrio cholerae*. *Infect Immun* 1995;63:4091–4098. [PubMed: 7558324]
19. Folster JP, Connell TD. The extracellular transport signal of the *Vibrio cholerae* endochitinase (ChiA) is a structural motif located between amino acids 75 and 555. *J Bacteriol* 2002;184:2225–2234. [PubMed: 11914354]
20. Mattick JS, Alm RA. Response from Mattick and Alm: common architecture of type 4 fimbriae and complexes involved in macromolecular traffic. *Trends Microbiol* 1995;3:411–413.
21. Filloux A, Michel G, Bally M. GSP-dependent protein secretion in gram-negative bacteria: the Xcp system of *Pseudomonas aeruginosa*. *FEMS Microbiol Rev* 1998;22:177–98. [PubMed: 9818381]
22. Sauvonnnet N, Vignon G, Pugsley AP, Gounon P. Pilus formation and protein secretion by the same machinery in *Escherichia coli*. *EMBO J* 2000;19:2221–2228. [PubMed: 10811613]
23. Sandkvist M. Biology of type II secretion. *Mol Microbiol* 2001;40:271–83. [PubMed: 11309111]
24. Winther-Larsen HC, Wolfgang M, Dunham S, van Putten JP, Dorward D, Lovold C, Aas FE, Koomey M. A conserved set of pilin-like molecules controls type IV pilus dynamics and organelle-associated functions in *Neisseria gonorrhoeae*. *Mol Microbiol* 2005;56:903–917. [PubMed: 15853879]

25. Peabody CR, Chung YJ, Yen MR, Vidal-Ingigliardi D, Pugsley AP, Saier MH Jr. Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella. *Microbiology* 2003;149:3051–3072. [PubMed: 14600218]
26. Mattick JS, Whitchurch CB, Alm RA. The molecular genetics of type-4 fimbriae in *Pseudomonas aeruginosa*--a review. *Gene* 1996;179:147–155. [PubMed: 8955641]
27. Wolfgang M, van Putten JP, Hayes SF, Dorward D, Koomey M. Components and dynamics of fiber formation define a ubiquitous biogenesis pathway for bacterial pili. *EMBO J* 2000;19:6408–6418. [PubMed: 11101514]
28. Marsh JW, Taylor RK. Identification of the *Vibrio cholerae* type 4 prepilin peptidase required for cholera toxin secretion and pilus formation. *Mol Microbiol* 1998;29:1481–1492. [PubMed: 9781884]
29. Aas FE, Winther-Larsen HC, Wolfgang M, Frye S, Lovold C, Roos N, van Putten JP, Koomey M. Substitutions in the N-terminal alpha helical spine of *Neisseria gonorrhoeae* pilin affect Type IV pilus assembly, dynamics and associated functions. *Mol Microbiol* 2007;63:69–85. [PubMed: 17140412]
30. Craig L, Taylor RK, Pique ME, Adair BD, Arvai AS, Singh M, Lloyd SJ, Shin DS, Getzoff ED, Yeager M, Forest KT, Tainer JA. Type IV pilin structure and assembly: X-ray and EM analyses of *Vibrio cholerae* toxin-coregulated pilus and *Pseudomonas aeruginosa* PAK pilin. *Mol Cell* 2003;11:1139–1150. [PubMed: 12769840]
31. Pasloske BL, Scraba DG, Paranchych W. Assembly of mutant pilins in *Pseudomonas aeruginosa*: formation of pili composed of heterologous subunits. *J Bacteriol* 1989;171:2142–2147. [PubMed: 2564847]
32. Burrows LL. Weapons of mass retraction. *Mol Microbiol* 2005;57:878–888. [PubMed: 16091031]
33. Nunn D. Bacterial type II protein export and pilus biogenesis: more than just homologies? *Trends Cell Biol* 1999;9:402–408. [PubMed: 10481178]
34. Durand E, Bernadac A, Ball G, Lazdunski A, Sturgis JN, Filloux A. Type II protein secretion in *Pseudomonas aeruginosa*: the pseudopilus is a multifibrillar and adhesive structure. *J Bacteriol* 2003;185:2749–2758. [PubMed: 12700254]
35. Nunn DN, Lory S. Cleavage, methylation, and localization of the *Pseudomonas aeruginosa* export proteins XcpT, -U, -V, and -W. *J Bacteriol* 1993;175:4375–4382. [PubMed: 8331069]
36. Durand E, Michel G, Voulhoux R, Kurner J, Bernadac A, Filloux A. XcpX controls biogenesis of the *Pseudomonas aeruginosa* XcpT-containing pseudopilus. *J Biol Chem* 2005;280:31378–31389. [PubMed: 16012171]
37. Kuo WW, Kuo HW, Cheng CC, Lai HL, Chen LY. Roles of the minor pseudopilins, XpsH, XpsI and XpsJ, in the formation of XpsG-containing pseudopilus in *Xanthomonas campestris* pv. *campestris*. *J Biomed Sci* 2005;12:587–599. [PubMed: 16078004]
38. Hu NT, Leu WM, Lee MS, Chen A, Chen SC, Song YL, Chen LY. XpsG, the major pseudopilin in *Xanthomonas campestris* pv. *campestris*, forms a pilus-like structure between cytoplasmic and outer membranes. *Biochem J* 2002;365:205–211. [PubMed: 11931643]
39. Vignon G, Kohler R, Larquet E, Giroux S, Prevost MC, Roux P, Pugsley AP. Type IV-like pili formed by the type II secretin: specificity, composition, bundling, polar localization, and surface presentation of peptides. *J Bacteriol* 2003;185:3416–3428. [PubMed: 12754241]
40. Lindeberg M, Salmond GP, Collmer A. Complementation of deletion mutations in a cloned functional cluster of *Erwinia chrysanthemi* out genes with *Erwinia carotovora* out homologues reveals OutC and OutD as candidate gatekeepers of species-specific secretion of proteins via the type II pathway. *Mol Microbiol* 1996;20:175–190. [PubMed: 8861215]
41. Jones S, Thornton JM. Protein-protein interactions: a review of protein dimer structures. *Prog Biophys Mol Biol* 1995;63:31–65. [PubMed: 7746868]
42. Jones S, Thornton JM. Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* 1996;93:13–20. [PubMed: 8552589]
43. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402. [PubMed: 9254694]

44. Lu B, Zhang D, McCammon JA. Computation of electrostatic forces between solvated molecules determined by the Poisson-Boltzmann equation using a boundary element method. *J Chem Phys* 2005;122:214102. [PubMed: 15974723]
45. DeLano, WL. The PyMOL Molecular Graphics System. 2002. www.pymol.org
46. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. *J Mol Biol* 1993;233:123–138. [PubMed: 8377180]
47. Hazes B, Sastry PA, Hayakawa K, Read RJ, Irvin RT. Crystal structure of *Pseudomonas aeruginosa* PAK pilin suggests a main-chain-dominated mode of receptor binding. *J Mol Biol* 2000;299:1005–1017. [PubMed: 10843854]
48. Keizer DW, Slupsky CM, Kalisiak M, Campbell AP, Crump MP, Sastry PA, Hazes B, Irvin RT, Sykes BD. Structure of a pilin monomer from *Pseudomonas aeruginosa*: implications for the assembly of pili. *J Biol Chem* 2001;276:24186–24193. [PubMed: 11294863]
49. Craig L, Volkmann N, Arvai AS, Pique ME, Yeager M, Egelman EH, Tainer JA. Type IV Pilus Structure by Cryo-Electron Microscopy and Crystallography: Implications for Pilus Assembly and Functions. *Mol Cell* 2006;23:651–662. [PubMed: 16949362]
50. Ramboarina S, Fernandes PJ, Daniell S, Islam S, Simpson P, Frankel G, Booy F, Donnenberg MS, Matthews S. Structure of the bundle-forming pilus from enteropathogenic *Escherichia coli*. *J Biol Chem* 2005;280:40252–60. [PubMed: 16172128]
51. Parge HE, Forest KT, Hickey MJ, Christensen DA, Getzoff ED, Tainer JA. Structure of the fibre-forming protein pilin at 2.6 Å resolution. *Nature* 1995;378:32–8. [PubMed: 7477282]
52. Craig L, Pique ME, Tainer JA. Type IV pilus structure and bacterial pathogenicity. *Nat Rev Microbiol* 2004;2:363–78. [PubMed: 15100690]
53. Forest KT, Dunham SA, Koomey M, Tainer JA. Crystallographic structure reveals phosphorylated pilin from *Neisseria*: phosphoserine sites modify type IV pilus surface chemistry and fibre morphology. *Mol Microbiol* 1999;31:743–752. [PubMed: 10048019]
54. Douet V, Loiseau L, Barras F, Py B. Systematic analysis, by the yeast two-hybrid, of protein interaction between components of the type II secretory machinery of *Erwinia chrysanthemi*. *Res Microbiol* 2004;155:71–75. [PubMed: 14990257]
55. O'Neal CJ, Jobling MG, Holmes RK, Hol WG. Structural basis for the activation of cholera toxin by human ARF6-GTP. *Science* 2005;309:1093–1096. [PubMed: 16099990]
56. Mosavi LK, Minor DL Jr, Peng ZY. Consensus-derived structural determinants of the ankyrin repeat motif. *Proc Natl Acad Sci U S A* 2002;99:16029–16034. [PubMed: 12461176]
57. Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter, C.; Sweet, R., editors. *Methods Enzymol.* 276. New York Academic Press; New York: 1997. p. 307–326.
58. Terwilliger T. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J Synchrotron Radiat* 2004;11:49–52. [PubMed: 14646132]
59. Morris RJ, Perrakis A, Lamzin VS. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol* 2003;374:229–244. [PubMed: 14696376]
60. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004;60:2126–2132. [PubMed: 15572765]
61. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 1997;53:240–255. [PubMed: 15299926]
62. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* 2006;62:439–450. [PubMed: 16552146]
63. Laskowski RA, Moss DS, Thornton JM. Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* 1993;231:1049–1067. [PubMed: 8515464]
64. Corpet F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 1988;16:108810890
65. Gouet P, Robert X, Courcelle E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* 2003;31:3320–3323. [PubMed: 12824317]

66. Martin J, Letellier G, Marin A, Taly JF, de Brevern AG, Gibrat JF. Protein secondary structure assignment revisited: a detailed analysis of different assignment methods. *BMC Struct Biol* 2005;5:17. [PubMed: 16164759]
67. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 2004;60:2256–2568. [PubMed: 15572779]
68. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* 2005;33:W299–302. [PubMed: 15980475]
69. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 1997;18:2714–23. [PubMed: 9504803]
70. Diederichs K, Karplus PA. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* 1997;4:269–75. [PubMed: 9095194]
71. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 2007;35:W375–83. [PubMed: 17452350]
72. Audette GF, Irvin RT, Hazes B. Crystallographic analysis of the *Pseudomonas aeruginosa* strain K122-4 monomeric pilin reveals a conserved receptor-binding architecture. *Biochemistry* 2004;43:11427–11435. [PubMed: 15350129]

Figure 1. The structure, topology and crystal contacts of EpsH

a) The structure of monomer A of EpsH³⁰⁻¹⁸⁸ with the corresponding topology diagram. Both the structure and topology diagram are colored according to the Type 4a pilin fold, with the conserved N-terminal α -helix in blue, the variable region with β -sheet I in purple and the conserved β -sheet II in green. A stretch of residues from Phe122 to Lys134 is disordered, and could not be included in the model, is denoted by a dotted black line. Chain B is very similar to Chain A, but is missing a larger part of the flexible loop (Gly104-Lys134).

b) Two molecules A and B of EpsH³⁰⁻¹⁸⁸ in one asymmetric unit (ASU) with neighboring ASU's on either side. Monomer A of EpsH³⁰⁻¹⁸⁸, denoted in blue by the letter A, forms three substantial contacts (labeled A-B, A-B' and A-B'') with three adjacent chains within the crystal lattice. Symmetry mates are labeled A' and B' for chains in the ASU to the left of monomer A, and A'' and B'' for chains to the right of Chain A.

Figure 2. Family sequence alignment of EpsH/GspH

A black arrow denotes the start of the construct for EpsH³⁰⁻¹⁸⁸. Secondary structure elements are shown on top and colored according to the Type 4a pilin fold, with the conserved N-terminal α -helix in blue, the variable region in purple and the conserved β -sheet labeled in green. Invariant residues are highlighted in red, while residues that are conserved in all but 1 or 2 homologues are highlighted in pink. A yellow background denotes medium sequence conservation and white denotes poor sequence conservation. Residues at positions that have a hydrophobic residue in all 10 homologues are indicated with red letters. The solvent accessibility as assigned by ESPRIPT on the basis of the EpsH³⁰⁻¹⁸⁸ structure is denoted by the bar *acc*, with dark blue meaning residues that are highly accessible, cyan meaning residues that are partially accessible and white denoting residues that are buried in the EpsH structure. Intermolecular contacts between A-B, A-B' and A-B'' are represented by triangles, which represent hydrophobic contacts, and by stars, which indicate residues that form H-bonds. Symbols for the contact residues in the A-B interface, (Gln37, Tyr40, Gln41, Leu44, Asn47, Glu48, Ile51, Leu52, Ly β 86, Asn174, Gly175, Thr176, Leu179) are colored in blue; in the A-B' interface, (Asp30, Ala32, Gln33, Pro158, Gly160, Gln161, Glu162, Asp164, Glu165, Gln166, Trp167, Ala181, Pro182, Gly183, Glu184, Ser185) are colored in green; and in the A-B'' interface, (Gly54, Gln55, Asp56, Leu73, Thr74, Ala75, Asp76, Gly78, Trp79, Phe142, Leu144, Ser145, Ser146, Glu148, Val149, Thr150, Pro151) are colored in pink. The black circles denote conserved hydrophobic residues that make up a conserved patch on the surface of EpsH.

Figure 3. Sequence conservation and electrostatics surface view of *Vibrio cholerae* EpsH

a) The sequence conservation as determined by Consurf⁶⁰ plotted on the surface of EpsH. Blue indicates high, red indicates low sequence conservation. A black arrow points to a highly conserved groove on the surface of EpsH, which is positioned between the variable and conserved β -sheets.

b) The electrostatic surface of EpsH³⁰⁻¹⁸⁸ as calculated by APBS tools⁴¹, in the same orientation as (a). A black arrow, which points to the highly conserved groove indicated in (a), shows that highly conserved residues form a hydrophobic crevice on the surface of EpsH.

c) A cluster of 17 conserved residues in stick representation that form the hydrophobic groove shown in figures 3a and 3b. These 17 residues, of which 15 are hydrophobic, may form a binding site for EpsH to interact with partner proteins or small molecules. Six residues that are invariant in all 10 homologues are labeled in red. Eight residues that conserve a hydrophobic residue are labeled in orange. Three residues that are highly conserved in 8-9 homologues are labeled in black. Four residues that are part of the interface 3 between monomers A and B'' in the crystal lattice are highlighted with a yellow background.

Figure 4. Comparison of *Vibrio cholerae* EpsH with all known Type 4a-like Pilin structures to date

a) EpsH is shown schematically along side the structures of equivalent structures of GspG from *K. oxytoca* (PDB code 1T92)¹⁵, PilA^{PAK} from *P. aeruginosa* (PDB code 1DZO)⁴⁸, PilE from *N. gonorrhoeae* (PDB code 2HI2)⁵³, and PilA^{K1224} from *P. aeruginosa* (PDB code 1RGO)⁷². To facilitate the comparison, all structures are shown in the same orientation of the conserved β -sheet colored in green. The fourth β -strand of *K. oxytoca* GspG is colored yellow to indicate a strand exchange between two equivalent β -strands of two chains observed in the asymmetric unit of *K. oxytoca* GspG¹⁵. These authors propose that the 4th strand under physiological conditions is provided by the same subunit as the three other strands.

b) Topology Diagrams of the three T4BP pilins and the two T2SS pseudopilins of known structure.

c) A structure-based alignment of 67 residues of EpsH and PilA^{PAK} that can be superimposed with an r.m.s.d. of 2.5 Å for 67 C α . The region that is most conserved between PilA^{PAK} and EpsH is indicated by a pink box. This corresponds to the same region that has the highest sequence conservation within the EpsH³⁰⁻¹⁸⁸ structure (Figure 2). Three residues that have

their numbers highlighted in red (Pro139 in EpsH, Gly175 in EpsH and Gly125 in PilA^{PAK}) are conserved within the EpsH and PilA families, respectively. The black arrows denote two *cis* prolines in PilA that are in the same location in EpsH but are in the *trans* conformation.

d) N-terminal alignment of EpsH and PilA^{PAK}. Shown are the 29 N-terminal residues in EpsH, and the 33 N-terminal residues in PilA^{PAK}, that were deleted from these proteins to obtain the respective crystal structures.

Figure 5. A possible position of EpsH at the tip of the T2SS pseudopilus

The conserved hydrophobic crevice of *Vibrio cholerae* EpsH facing outwards when EpsH is placed at the tip of a tentative pseudopilus model (for details see Materials and Methods). The model T2SS pseudopilus was constructed from EpsG homology-modeled subunits, based on the *K. oxytoca* GspG structure¹⁵, that are superimposed onto PilE subunits within the assembled PilE pilus⁴⁹. The modeled EpsG subunits are colored dark green. The helix of PilE is colored light green. EpsH superimposed onto the tip of the PilE pilus is colored red. The 18 residues that form the hydrophobic crevice on the surface of EpsH are colored yellow.

Table 1
Data collection and refinement statistics *V. cholerae* EpsH

<i>Data collection and processing</i>	
Wavelength (Å)	0.97903
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁
<i>a</i> (Å)	53.3
<i>b</i> (Å)	70.4
<i>c</i> (Å)	85.1
Resolution (Å)	20-2.00 (2.07-2.00)
Unique reflections	20742
Completeness (%)	93.1 (100.0)
Redundancy	7.1 (7.1)
R_{meas} (%) ^a	7.4 (28.8)
R_{sym} (%) ^b	6.9 (26.7)
<i>I</i> /σ(<i>I</i>)	18.5 (7.4)
Wilson <i>B</i> -factor (Å ²)	21.6
<i>Refinement</i>	
Atoms	2525
Molecules per ASU	2
Residues per ASU	268
Waters per ASU	301
R_{work} (%)	16.9
R_{free} (%)	21.2
RMSD bond lengths (Å)	0.016
RMSD bond angles (°)	1.54
RMSD chiral centers (Å ³)	0.094
<i>B</i> average EpsH (Å ²)	23.8
<i>B</i> average water (Å ²)	37.4
% Residues in favored Ramachandran region	99.6%
% Residues in allowed Ramachandran region	0.4%

The data collection statistics were extracted from HKL2000⁵⁷ and NovelR⁷⁰. The refinement statistics were calculated by REFMAC⁶¹. Data in parentheses refer to the highest resolution shell. The Ramachandran statistics are from Molprobity server⁷¹. Molprobity: all-atom contacts and structure validation for proteins and nucleic acids⁷¹. The model *B*-factors are reported with TLS component included.

^a $R_{\text{meas}} = (\sum [N/(N-1)])^{1/2} \sum |I - \langle I \rangle| / \sum I$, where *N* is the redundancy of a given reflection⁷⁰.

^b $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum | \langle I \rangle |$, where *I* is the integrated intensity of a given reflection.