Published in final edited form as:

J Mol Biol. 2012 May 25; 419(1-2): 110–124. doi:10.1016/j.jmb.2012.02.041.

Structure of an Essential Type IV Pilus Biogenesis Protein Provides Insights into Pilus and Type II Secretion Systems

Atsushi Yamagata^{a,†}, Ekaterina Milgotina^{b,†}, Karen Scanlon^b, Lisa Craig^c, John A. Tainer^{a,*}, and Michael S. Donnenberg^{b,*}

^aThe Scripps Research Institute Molecular Biology, MB4, Skaggs Institute for Chemical Biology, 10550 North Torrey Pines Rd, La Jolla, CA92037

^bDivision of Infectious Diseases, Department of Medicine, University of Maryland School of Medicine, 20 Penn Street, Baltimore, MD 21201

^cDepartment of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC Canada, V5S 1S6

Abstract

Type IV pili (T4Ps) are long cell surface filaments, essential for microcolony formation, tissue adherence, motility, transformation, and virulence by human pathogens. The enteropathogenic *E. coli* bundle-forming pilus (BFP) is a prototypic T4P assembled and powered by BfpD, a conserved GspE secretion superfamily ATPase held by inner membrane proteins BfpC and BfpE, a GspF-family membrane protein. Although the T4P assembly machinery shares similarity with type II secretion (T2S) systems, the structural biochemistry of the T4P machine has been obscure. Here, we report the crystal structure of the two-domain BfpC cytoplasmic region (N-BfpC), responsible for binding to ATPase BfpD and membrane protein BfpE. The N-BfpC structure reveals a prominent central cleft between two α/β domains. Despite negligible sequence similarity, N-BfpC resembles PilM, a cytoplasmic T4P biogenesis protein. Yet surprisingly, N-BfpC has far greater structural similarity to T2S component EpsL, with which it also shares virtually no sequence identity. The C-terminus of the cytoplasmic domain, which leads to the transmembrane segment not present in the crystal structure, exits N-BfpC at a positively-charged surface that most likely interacts with the inner membrane, positioning its central cleft for interactions with other Bfp components. Point mutations in surface-exposed N-BfpC residues predicted to be critical for interactions among BfpC, BfpE and BfpD disrupt pilus biogenesis without precluding interactions with BfpE and BfpD and without affecting BfpD ATPase activity. These results illuminate the relationships between T4P biogenesis and T2S systems, imply that subtle changes in component residue interactions can have profound effects on function and pathogenesis, and suggest that T4P systems may be disrupted by inhibitors that do not preclude component assembly.

© 2012 Elsevier Ltd. All rights reserved

*Corresponding authors. Phone: +1 410-706-7562, Fax, +1 410-706-7562, mdonnenb@umaryland.edu ; JATainer@lbl.gov.

†Contributed equally

‡Contributed equally

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.

Accession number The atomic coordinates of N-BfpC are deposited in the Protein Data Bank, PDB ID 3VHJ.

Keywords

Type IV pili; Type II secretion; crystal structure; GspE ATPase; protein-protein interaction

Introduction

Type IV pili (T4Ps)¹ or fimbriae are surface appendages by numerous species of Gram-negative produced bacteria, some Gram-positive bacteria and some Archaea.^{1, 2} Among these microorganisms are a plethora of important pathogens responsible for a massive burden of human illness. These diseases include infections of the upper respiratory tract (*Haemophilus influenzae*, *Moraxella catarrhalis*), pneumonia and other infections in compromised hosts (*Legionella pneumophila*, *Pseudomonas aeruginosa*), meningitis (*Neisseria meningitidis*), sexually transmitted infections (*N. gonorrhoeae*), diarrhea (*Vibrio cholerae*, enteropathogenic *Escherichia coli* (EPEC), enterotoxigenic *E. coli*, Shiga-toxin producing *E. coli*, *Yersinia pseudotuberculosis*), enteric fever (*Salmonella enterica* serovar Typhi), and other systemic infections (*Burkholderia pseudomallei*, *V. vulnificus*). Numerous functions are ascribed to T4Ps including adherence to tissue culture cells, auto-aggregation, twitching motility, and formation of biofilms.¹ Several T4Ps have been confirmed in animal models^{3–6} or volunteer studies to be required for colonization or virulence.^{7–9} Thus, a thorough understanding of the biology of T4Ps has important implications for microbial biology and human health.

T4Ps are composed of repeating subunits of pilin, which are synthesized as pre-pilin precursors.¹ Pre-pilin is processed by the pre-pilin peptidase, a bi-functional aspartic acid protease that removes the signal sequence and *N*-methylates the terminal residue of the mature protein.^{10, 11} The periplasmic oxido-reductase DsbA catalyzes the formation of a disulfide bond in the globular domain of the pre-pilin.¹² As the active site of the pre-pilin peptidase is in the cytoplasm and cleavage of the signal sequence and DsbA-catalyzed disulfide bond formation occur simultaneously and independently, the pre-pilin is an integral cytoplasmic transmembrane protein with the N-terminus in the cytoplasm, a conserved hydrophobic α -helical transmembrane domain, and a globular periplasmic domain. Crystal structures of full-length pilin subunits and electron microscopy analysis of the pilus filament reveal that the conserved hydrophobic N-terminal sequence forms a long α -helix that is bundled in the core of the mature T4P filament.^{13–15} A complex multimeric T4P biogenesis machine is required for extraction of the pilin protein from the cytoplasmic membrane, assembly of the pilin monomers into helical pili, and extrusion of the pili through the outer membrane (OM).¹⁶

T4Ps have been divided into two groups with evidence of divergent evolutionary histories.¹⁷ The T4aP systems are distinguished by short pre-pilin leader peptides, smaller pilin proteins, and the scattering of genes encoding essential proteins at various sites in the genome. The genes encoding T4bP systems are found on a single genetic locus, the pre-pilin leader peptides are longer and the pilin proteins larger. The bundle-forming pilus (BFP) of EPEC, a T4bP, has emerged as an important prototypic system to study T4P biogenesis and function. The fourteen genes of the *bfp* operon are sufficient to allow BFP assembly when expressed from an artificial promoter in a laboratory *E. coli* host.¹⁸ The *bfpA* gene encodes the pre-pilin protein, pre-bundlin,¹⁹ while the *bfpP* gene encodes the pre-pilin

¹Abbreviations used: BFP, bundle-forming pilus; EPEC, enteropathogenic *Escherichia coli*; IM, inner membrane; IPTG, isopropyl β -D-1-thiogalactopyranoside; MAD, multi-wavelength anomalous dispersion; N-BfpC, cytoplasmic amino terminal region of BfpC; Ni-NTA, nickel nitrilotriacetic acid; OM, outer membrane; PMSF, phenylmethanesulfonyl fluoride; RMSD, root-mean-square distance; Se-Met, selenomethionine; T2S, type II secretion; T4P, type IV pilus; T4aP, type IVa pilus; T4bP, type IVb pilus.

peptidase.²⁰ Aside from *bfpH*, which seems to be a pseudogene,^{16, 18} and *bfpF*, which encodes a member of the PilT family ATPases required for disassembly of auto-aggregates (disaggregation) and retraction of the pili,^{7, 21} the other ten genes encode proteins required for pilus assembly.^{22, 23} Ten of the Bfp proteins including bundlin and BfpF, but excluding BfpP and two other pre-pilin peptidase substrates (also known as pseudo-pilins) BfpI and BfpK, can be detected after cross-linking and purification of the OM secretin BfpB, suggesting that they form a complex spanning both the inner and OM.¹⁶

To facilitate analysis of the BFP biogenesis machine, we conceptually divide the Bfp proteins into components of inner membrane (IM) and OM subassemblies. The IM subassembly is composed of the two cytoplasmic ATPases BfpD and BfpF plus two transmembrane proteins, BfpC and BfpE.²⁴ BfpD is a member of the highly conserved PilB/GspE hexameric secretion ATPase superfamily, whose members act in many bacterial and archaeal macromolecular transport systems, such as Type II secretion (T2S), type IV secretion, DNA uptake by natural transformation and conjugation, and archaeal flagellar assembly systems.²⁵ BfpE is a member of the GspF secretion family of membrane proteins, which are conserved in T4P, T2S, and Gram-positive competence systems, with recognizable orthologs in archaeal flagellar systems.²⁶ A detailed topological analysis suggested that BfpE has four transmembrane segments with a 114-residue cytoplasmic N-terminal domain, N-BfpE.²⁷ However, *in silico* analysis suggests that the proposed second periplasmic domain is homologous to the N-terminal domain and the two domains may associate in the cytoplasm.²⁸ BfpC is a bitopic IM protein with the N-terminal residues 1–164 in the cytoplasm (N-BfpC) and C-terminal residues 188–402 in the periplasm.²⁴ Together BfpC and BfpE recruit BfpD to the membrane, while genetic evidence suggests that BfpF interacts with a small cytoplasmic sequence of BfpE.²⁴

T4P biogenesis systems are closely related to T2S systems. Indeed, the T4aP system of *P. aeruginosa* is capable of assembling components of T2S systems into pili²⁹ and over expression of a T2S system can lead to formation of pili.³⁰ Moreover, some T4P systems have secretory functions.^{31, 32} These observations have led to the hypothesis that T2S systems use reversible assembly of pilin-like proteins into short filaments as a piston to pump substrates across the outer membrane. The GspE T2S ATPase forms a tight complex with the cytoplasmic domain of the GspL family bitopic membrane protein, as shown by the crystal structure of the complex between EpsE and EpsL from *V. cholerae*.^{33, 34} However, limited structure-based information about T4P assembly machines has restricted our understanding of their function. To define the molecular architecture of a prototypic T4P machine and address questions regarding the membrane-associated components of T4P and their relationship to the T2S system, we purified, crystallized and solved the crystal structure of N-BfpC and analyzed the structure for insights into the T4P machine and related T2S system.

Results

N-BfpC crystal structure, architecture and electrostatic surface

To help define the structural basis for BfpC interactions acting in the T4P IM subassembly, we expressed, purified and crystallized the cytoplasmic region of the protein (residues 1–164, N-BfpC). We solved the crystal structure of N-BfpC by multi-wavelength anomalous dispersion (MAD) phasing and refined the resulting structure to 1.9 Å resolution (Table 1). Residues 3–164 were visible in electron density maps. N-BfpC, which crystallized as a monomer, is a globular protein with a distinct cleft separating two similar domains, each comprised of α -helix- β -sheet combination (Figure 1A). The protein begins with an α -hairpin (strands β 1 and β 2), which lies between the two domains at the back of the protein opposite to the cleft. The β -hairpin is followed by a short β -strand, β 3, the first of three anti-parallel

strands for the domain I β -sheet. A short loop connects $\beta 3$ to an α -helix, $\alpha 1$, which runs parallel to the β -strands. Strands $\beta 4$ and $\beta 5$ follow, after which comes an irregular loop that runs along the back of the protein from the top of domain I to the bottom of domain II. A single α -helical turn is located in the center of this extended loop. The loop feeds into antiparallel strands $\beta 6$ – $\beta 8$ of the 4-stranded domain II β -sheet. $\beta 8$ is followed by $\alpha 2$, which, like $\alpha 1$, runs parallel to its associated β -strands. From $\alpha 2$, the polypeptide chain forms the last β -strand, $\beta 9$. This short β -strand is adjacent and parallel to $\beta 6$ on the outside of the domain II β -sheet. Thus, while the overall structures of domains I and II are similar they have distinct topologies. From $\beta 9$, the polypeptide chain forms an irregular loop that again wraps around the back of the domain to lie across domain I. A 3_{10} helix is located at the center of this terminal extended loop. The two helical turns and the β -hairpin form a sub-domain at the back of the protein, comprised of N- and C-terminal segments and flanked by the β -sheets. The cleft at the front of the protein separates domains I and II. The peripheral β -strands ($\beta 3$ and $\beta 8$) of the two β -sheets run parallel to each other and form the floor of the cleft but do not form canonical β -sheet hydrogen bonds. The cleft is framed by the two α -helices and lined with several exposed hydrophobic amino acids as well as polar and charged residues.

The C-terminus of N-BfpC, Lys164, is followed in native BfpC by the transmembrane segment residues 165–187.³⁵ The region surrounding Lys164 is rich in basic residues that form a strong positively charged patch based on its electrostatic surface potential (Figure 1B). This surface likely interacts with the negatively charged phospholipid head groups of the inner membrane, which would expose the central cleft of BfpC to the cytoplasm (Figure 1C).

Structural similarity of N-BfpC with T2S inner membrane protein EpsL and cytoplasmic T4aP protein PilM

As N-BfpC has no sequence homologues among known proteins, we performed a comprehensive search for structural homologues using the DALI server.³⁶ N-BfpC is similar to the N-terminal domain of EpsL from *V. cholerae* (N-EpsL, PDB ID 1YF5, DALI Z-score of 5.2). EpsL is an IM protein that binds to the T2S ATPase EpsE.^{33, 37} We showed previously that BfpC and BfpE recruit the BFP assembly ATPase BfpD to the membrane.²⁴ Thus, BfpC and EpsL are functional and structural homologs, despite the absence of recognizable sequence similarity between these proteins (Figure 2A).²⁴ The globular N-EpsL has 3 $\alpha\beta$ -domains and domains II and III superimpose well on domains I and II of N-BfpC (root-mean-square deviation, RMSD, 3.3 Å for C α atoms of 132 aligned residues), indicating significant structural similarity (Figure 2B). The topology of the corresponding domains is identical in the two proteins. Domain I of N-EpsL lies at the back of the protein behind the cleft but is larger and more complex than the sub-domain of N-BfpC, with 3 α -helices enveloping a central β -sheet.

In the crystal structure of a complex between N-EpsL and the N-terminus of the T2S ATPase N1-EpsE (PDB ID 2BH1), the $\alpha 2$ α -helix of N1-EpsE is inserted into the top of the N-EpsL cleft and the short N-terminal N1-EpsE α -helix, $\alpha 1$, interacts with a loop between the β -sheet and the α -helix of N-EpsL domain II (Figure 2C). In total, N-EpsL interacts with N1-EpsE via three discrete regions (Figure 2C, magenta): (1) residues 61–65; (2) 87–91; and (3) 168–181. N-EpsL region 1 residues D61, L62, and I63, the backbones of which interact with N1-EpsE, correspond to residues N19, L20, and L21 in N-BfpC, respectively (region 1, Figure 2D). N-EpsL region 3 side chains of M168, W174, L177 and L178 form hydrophobic interactions with N1-EpsE. In the corresponding site in N-BfpC, K111, F122 and V127 are oriented into the cleft (region 2, Figure 2D). Thus, these exposed cleft residues in N-BfpC may be involved in binding to the N-terminus of the EPEC assembly ATPase BfpD in a manner similar to the N-EpsL:N1-EpsE interaction. Both N-EpsL alone, and its complex

with N1-EpsE formed dimers in their crystal structures, although N-EpsL does not form a dimer in solution.³³ On the other hand, N-BfpC is a monomer in both solution and crystal structure. The N-BfpC structure lacks the region corresponding to the domain I of N-EpsL, which is responsible for its dimer interface, thus explaining the different oligomerization states in the respective crystal structures.

The structure of the cytoplasmic PilM T4aP biogenesis protein from *Thermophilus* (PDB ID 2YCH) also resembles N-BfpC (Figure 2B) in spite of the lack of detectable sequence similarity between the two proteins.³⁸ Interestingly, the structure of PilM is far more similar to that of EpsL (reported RMSD of aligned C α atoms of 2.8 Å) than it is to N-BfpC (RMSD of 110 aligned C α atoms of 4.3 Å), despite the fact that PilM and N-BfpC are both from T4P systems while EpsL is from a T2S system. Domain I of EpsL, lacking in N-BfpC, is present in PilM. In addition PilM has a complete ATP-binding domain also lacking in N-BfpC, confirming and extending the previously noted similarity between EpsL and FtsA.³⁷ PilM was crystallized along with ATP and a peptide from the small cytoplasmic N-terminal segment of the bitopic T4P biogenesis protein PilN (Figure 2E). Interestingly, the PilN peptide occupies a cleft between domains I and II in PilM that in N-EpsL is occupied by its own C-terminal strand, which is immediately followed by the transmembrane domain in native EpsL. In N-BfpC there is no such cleft due to the absence of an equivalent of the EpsL domain I, however, its C-terminal segment nonetheless occupies a similar position along the back of domain I. This segment is part of the positively-charged patch on N-BfpC, predicted to interact with the inner membrane (Figure 1). Thus, whereas BfpC and EpsL are anchored to the inner membrane via their own transmembrane segments, PilM appears to use PilN for membrane anchorage.

Essential N-BfpC residues predicted to interact with BfpD

Given the structural similarities between N-BfpC and EpsL, we investigated whether residues in the BfpC cleft, particularly those corresponding to regions 1 and 3 of EpsL, are essential for BFP biogenesis.²⁴ Like EpsE, the purified BfpD assembly ATPase assembles into a hexamer (Supplemental Table 1, Supplemental Figure 1), a consistent structural feature of secretion ATPases.^{39, 40} BfpD may bind to the BfpC cleft in a manner analogous to the N1-EpsE and N-EpsL interaction.³³ Accordingly, N19, K111 and L21, which all lie on the floor of the cleft (Figure 3A), are candidates for mediating interactions between N-BfpC and BfpD. Residues T28, K37 and M42 are exposed on α 1 in domain I and thus may also participate in binding to partner proteins.

To test these predictions, we used site-directed mutagenesis on a plasmid (pEM26) designed to express BfpC variants in which residues of interest were replaced with alanine and N19 and K111 were replaced with aspartic and glutamic acid, respectively (Table 2). We assessed the ability of these plasmids to restore auto-aggregation and disaggregation to the EPEC *bfpC* mutant strain UMD924. The auto-aggregation phenotype is dependent on expression of BFP filaments and acts as a test for a functional assembly apparatus while disaggregation is dependent on BfpF and its putative pilus retraction function.²³ Importantly, immunoblotting revealed that each of the BfpC variants was expressed in the *bfpC* mutant background strain (Figure 3B). Plasmids pEM26, pKMS01, pEM92, pEM95, and pEM96, encoding wild-type BfpC, BfpC_{N19A}, BfpC_{N19D}, BfpC_{T28A}, and BfpC_{K37A}, respectively, were able to complement the *bfpC* mutant strain to restore auto-aggregation and disaggregation (Table 2). In contrast, plasmids pEM93, pEM94, pEM97, and pEM98, encoding BfpC_{L21A}, BfpC_{N19A,L21A}, BfpC_{M42A}, and BfpC_{K111E}, respectively, could not. Of note, the auto-aggregates formed by the strain containing pEM92 encoding BfpC_{N19D} were considerably smaller than those formed by the strains expressing wild type BfpC. To determine whether any *bfpC* mutations able to complement the *bfpC* mutant might have more subtle effects on pilus function, we calculated the aggregation index for each of these

strains. Figure 3C shows the data at 4 hours, a time point at which differences were maximized, which confirm the impressions obtained by visual inspection of auto-aggregates. As a further test of function, we used flow cytometry to measure surface bundlin immunofluorescence (Figure 3D). We found that strains expressing BfpC variants that preclude auto-aggregation have levels of surface bundlin immunofluorescence no greater than background seen in the *bfpC* mutant UMD924, while all those that permit auto-aggregation have more surface bundlin. These results reveal that residues L21, M42 and K111 are required for function of the BFP biogenesis machine, while N19, T28 and K37 are tolerant to mutation. Interestingly, while the N19A substitution did not preclude normal function of the BFP machinery and in fact resulted in increased levels of surface bundlin and auto-aggregation, mutation N19D reduced and K111E abolished the formation of auto-aggregates and surface bundlin expression. K111E and N19 are in very close proximity and along with L21 and M42 form a band along the top of the BfpC cleft. These residues are thus likely to make contact with any protein bound in this region.

Mutations that disrupt BFP biogenesis are insufficient to disrupt interactions among N-BfpC, BfpD and N-BfpE or to reduce BfpD ATPase activity

BfpC and BfpE together recruit the BfpD ATPase to the cytoplasmic membrane.²⁴ To assess the effect of *bfpC* mutations that interfere with function on interactions among these three proteins, we created a series of plasmids with an artificial operon from which we could express BfpD alone, BfpD with BfpC, or BfpD with both BfpC and BfpE. We separated lysates of *E. coli* strains carrying these plasmids into soluble and insoluble (membrane) fractions and detected BfpD by immunoblotting. GroEL, which served as a control for a cytoplasmic protein, was consistently present overwhelmingly in the soluble fraction (not shown). As previously reported,²⁴ BfpD was predominantly soluble when expressed alone, and insoluble when expressed in conjunction with BfpC and BfpE (Figure 4A), consistent with a role for BfpC and/or BfpE in recruiting BfpD to the membrane. When co-expressed only with wild type BfpC, BfpD remained predominantly soluble. Furthermore, co-expression of BfpD with *bfpC* mutations that interfere with autoaggregation did not substantially affect BfpD solubility. Similarly, when we expressed BfpD with both BfpC and BfpE, we found no substantial effect of *bfpC* mutations that preclude BFP biogenesis on BfpD solubility (Figure 4B). Thus, amino acid changes in BfpC that disrupt pilus function do not appear to alter the ability of the inner membrane proteins to recruit BfpD.

Next, we used co-affinity purification to determine the effect of amino acid substitutions in N-BfpC on the stability of the *in vitro* interactions among N-BfpC, N-BfpE and BfpD. We attached a hexahistidine tag to the C-terminus of BfpD and created a series of plasmids with an artificial operon encoding wild type or mutant N-BfpC, wild type N-BfpE, and BfpD-His. As a control we also created a plasmid with wild type N-BfpC and N-BfpE without BfpD. We applied lysates from strains containing these plasmids to a nickel affinity column, washed, and eluted bound proteins with imidazole. Coomassie blue staining revealed BfpD-His in all eluate samples, with N-BfpC and N-BfpE visible as very faint bands, if at all, and no other visible proteins (Supplemental Figure 2). We performed immunoblotting for BfpC and BfpE and quantified the amount of each protein recovered in the eluate as a percentage of the total amount in the flow-through, wash and eluate. Although there was considerable variation from experiment to experiment, abundant N-BfpC and N-BfpE could be detected in immunoblots after coelution from a nickel agarose column in the presence of BfpD-His, but significantly less of each protein was detected in the absence of BfpD (Figure 4C). However, co-elution of either N-BfpC or N-BfpE with BfpD-His was not significantly reduced by either the *bfpC*_{M42A} or the *bfpC*_{N19A L21A} mutation. The percentage of BfpD-His eluted from the column did not differ substantially among the various samples. These results indicate that mutations in BfpC that interfere with BFP assembly and are predicted to

interfere with specific interactions between BfpC and BfpD are insufficient to disrupt the BfpC-BfpD-BfpE complex.

The presence of partner proteins can induce allosteric alterations in enzyme activity for ATPases closely related to BfpD, such as EpsE.⁴¹ Therefore we decided to test the hypothesis that mutations in *bfpC* interfere with the function of the BFP biogenesis machine by reducing BfpD activity. Control experiments indicated that purified BfpD with the hexahistidine C-terminal tag has modest ATPase activity in the presence of magnesium and that this activity is increased approximately 3-fold when BfpD is purified along with N-BfpC and N-BfpE (Supplemental Figure 3). Therefore, we assayed the samples obtained by co-affinity purification of BfpD with wild type and variant N-BfpC for ATPase activity. We found no evidence that samples containing non-functional variant N-BfpC proteins exhibit reduced ATPase activity in comparison to the sample that had the wild type forms of these components (Figure 4D). The modest increase in activity observed is likely due to slightly lower total protein concentrations in the samples containing variant proteins.

Reciprocal mutations in N-BfpC and N-BfpE preserve function

Since we were unable to detect an effect of mutating BfpC residues predicted to interfere with BfpC-BfpD interactions on BfpD recruitment to the membrane, BfpD co-purification or BfpD function, we considered the possibility these amino acid changes might instead disrupt interactions with the N-terminus of BfpE. BfpE, which is a member of the GspF family of inner membrane proteins that is highly conserved in T4P, T2S and other related systems, might bind within the N-BfpC cleft in addition to or instead of BfpD. In support of this idea, a crystal structure of the N-terminal domain of the *V. cholerae* T2S GspF family member EpsF was solved as a dimer and was shown to be located on the cytoplasmic side of the inner membrane.²⁸ Alignment analysis of the N-terminal amino acid sequences of EpsF and BfpE revealed a highly conserved region corresponding to the $\alpha 5$ helix in EpsF, with 9 out of 16 residues being identical.²⁸ $\alpha 5$ is exposed on the surface of the EpsF dimer and thus may interact with partner proteins. Interestingly, two conserved residues, D130 and E131, on the N-terminus of the $\alpha 5$ helix in EpsF (D82 and E83 in BfpE) are exposed to the solvent and are predicted to lie in close proximity to the cytoplasmic membrane. Thus, these negatively charged amino acids, which would be expected to be repulsed by the negatively-charged phospholipid head groups, are candidates for critical interactions with N-BfpC polar residues such as the essential cleft residue K111 and/or N19. Thus, we sought to test the hypothesis that interactions between conserved residues D82 and E83 of BfpE with N19 and K111 of BfpC are required for T4P biogenesis.

Plasmid pKDS302 contains the entire *bfp* operon under the control of an isopropyl β -D-1-thiogalactopyranoside (IPTG)-inducible promoter and can direct the expression of BFPs in a laboratory *E. coli* host strain.¹⁸ In the presence of a constitutive *cpxA** mutation, sufficient pili are expressed to allow auto-aggregation.⁴² Because it is contained in a single plasmid of moderate size, this system is more suitable to engineering mutations in different *bfp* genes than is the system in an EPEC background strain. We created six modified plasmids from pKDS302 by site-directed mutagenesis with the following substitutions in *bfpC* and/or *bfpE*: *bfpC*_{N19D}, *bfpE*_{D82N}, *bfpC*_{N19D} and *bfpE*_{D82N}, *bfpC*_{K111E}, *bfpE*_{E83K}, and *bfpC*_{K111E} and *bfpE*_{E83K}. Although *E. coli* (strain ALN92) transformed with pKDS302 expressing wild-type BFP proteins exhibited auto-aggregation (Fig 4E), the same strain harboring pKDS302 derivatives with *bfpC*_{K111E}, *bfpE*_{D82N}, *bfpE*_{E83K}, or both *bfpC*_{K111E} and *bfpE*_{E83K} failed to exhibit auto-aggregation (Table 2). These results suggest that BfpE residues D82 and E83 are critical for pilus biogenesis and/or pilus functions. As expected from the complementation assay above, *E. coli* with pKDS302 carrying the *bfpC*_{N19D} mutation formed aggregates. Of note, the auto-aggregation assay could not be used to quantify these results as the aggregates

of even the strain carrying the plasmid with the wild type genes are too small to detect in this assay.

Significantly, while the *bfpE_{D82N}* mutation disrupted BFP biogenesis in the setting of wild type BfpC, *E. coli* carrying the pKDS302 derivative with dual *bfpC_{N19D}* and *bfpE_{D82N}* substitutions was capable of auto-aggregation. This extragenic suppression of a *bfpE* mutation that interferes with BFP assembly suggests that BfpC and BfpE might interact through a hydrogen bond between N19 of BfpC and D82 of BfpE, although more complex interactions involving these and other residues cannot be excluded. Furthermore, this experiment confirmed that K111 of BfpC is required for BFP biogenesis and determined that E83 of BfpE is also essential. Nevertheless, we were unable to detect an effect of the *bfpE_{D82N}* mutation on the ability to co-purify BfpE or BfpC with BfpD (Figure 4C) or on BfpD activity (Figure 4D).

Discussion

We determined the crystal structure of N-BfpC, which revealed that, despite negligible sequence similarities, N-BfpC is structurally related to the T2S component NEpsL and the cytoplasmic T4aP protein PilM, suggesting that these proteins share a common evolutionary origin and likely a similar function. However, key differences are also apparent. Obvious differences among these proteins concern their size and topology. BfpC is a bitopic IM protein with a 214-residue periplasmic domain (BfpC-C).²⁴ EpsL has a similar topology, but the periplasmic region has only 113 amino acids,⁴³ while PilM is entirely cytoplasmic.³⁸ In addition to its association with the EpsE ATPase, EpsL interacts directly with EpsM, which is another bitopic IM protein but one in which the cytoplasmic region is short with a larger periplasmic region.⁴³ In contrast, PilM interacts in a complex with PilN and PilO, which have topologies similar to EpsM, and with the predicted lipoprotein PilP.⁴⁴ The partial structures of EpsM, PilO and the periplasmic region of EpsL reveal that they are homologous proteins with common $\alpha\beta\beta$ circular permutations of the ferredoxin fold.^{45–47} It is tempting to predict that BfpC-C will prove to be similar, despite its greater size and negligible sequence homology. Indeed, secondary structure prediction algorithms suggest that BfpC-C has a more complicated arrangement that could include such a fold (data not shown).

The structure of N-BfpC reveals other differences from N-EpsL and PilM. For example, N-EpsL domain I, which is also present in PilM, is absent from N-BfpC. The function of this domain is not yet defined. The fact that N-EpsL forms a dimer in the crystal structure, while N-BfpC does not, highlights another difference between N-BfpC and N-EpsL. EpsM and the periplasmic region of EpsL each form homodimers and interact with each other.⁴⁵ Similarly, PilO forms homodimers and heterodimers with PilN.⁴⁷ Each of these proteins forms complexes with the structural homologues of N-BfpC, which suggests that BfpC might also dimerize in the presence of other Bfp proteins. Alternatively, the dimers of N-EpsL seen in the crystal structure may be an artifact of high protein concentration, since N-EpsL exists as monomer in solution, and N-EpsL/N1-EpsE exists as a heterodimer with one copy of each partner.³³

An additional functional distinction between N-EpsL and both N-BfpC and PilM is the presence of the positively charged patch on N-BfpC, which we predict to interact with membrane phospholipids. This positively charged patch is composed of the residues K45, K63, K68 and R69 from the loop following strand β 3 in domain I, and K157, K159, K160 and R162 from the long loop of the C-terminal end. When N-BfpC and N-EpsL are superimposed, the N-EpsL C-terminal end and part of domain I (α 2) correspond to this area. Thus, we predict that N-EpsL C-terminal residues 243–253, which were eliminated from the construct used for crystallization, play a key role in the interaction between EpsL-EpsE

complex and acidic phospholipids.⁴¹ Both the positively charged patch of N-BfpC and the corresponding region of N-EpsL may therefore participate in the interaction with phospholipid to analogously position both proteins with regard to the IM.

One of the most intriguing ramifications of these structural comparisons concerns the evolutionary relationships among T4aP, T4bP and T2S systems. Homologues of PilM proteins are present throughout T4aP biogenesis systems and are arranged in operons along with PilN, PilO, PilP and the secretin PilQ.⁴⁸ In contrast, homologues of these proteins, with the exception of the secretin, are not obvious among the T4bP systems. The N-BfpC structure reported herein reveals its relationship to PilM despite the lack of sequence similarity. This relationship stands in stark contrast to those of “core” components of T4P biogenesis systems including the pilins, pilin-like proteins, pre-pilin peptidase, secretin, ATPases, and polytopic inner membrane protein, all of which display obvious similarity among all such systems in Gram negative bacteria. One implication concerning the function of the PilM/BfpC protein is that its sequence may have been less constrained by evolution than that of the core components or that evolution favored diversification of this component. It is particularly intriguing that N-BfpC more closely resembles the T2S component N-EpsL in both structure and topology, than it does the T4aP component PilM. This paradox raises the possibility that T4bP systems may share a more recent common ancestor with T2S systems than they do with T4aP systems.

Inspection of the N-BfpC structure, and comparison with that of N-EpsL, identified several residues that potentially participate in critical interactions among N-BfpC, the BfpD ATPase, and N-BfpE, the amino terminus of the polytopic transmembrane protein BfpE. Mutational analysis revealed that some of these residues are essential for function of the BFP machine, as mutations in L21, M42 and K111 disrupt auto-aggregation and thus BFP biogenesis (Table 2). In contrast other residues, such as T28 and K37, both of which extend into the solvent, could be modified without disturbing function. Surprisingly, asparagine 19, which is positioned across the cleft from K111, and is analogous to D61 in EpsL, could be changed, not only to aspartic acid, but also to alanine and still retain function.

While certain BfpC residues that potentially participate in interactions with partner proteins are essential for function of the BFP machine, an understanding of the precise functional role for these residues remains elusive. For example, by analogy with the N-EpsL/N1-EpsE structure, L21 is predicted to be part of hydrophobic patch required for the interaction of BfpC with BfpD. However, changing this residue from leucine to alanine has no apparent effect on the ability of BfpC and BfpE to recruit BfpD to the membrane, even though it renders the BFP machine non-functional. Similarly this mutation, even in combination with an alteration of residue 19 from asparagine to aspartic acid, which seems to reduce BfpC function, had no effect on the co-purification of N-BfpC or N-BfpE with histidine-tagged BfpD and did not reduce the ATPase activity of BfpD in such preparations. Similar results were obtained when methionine residue 42 was changed to alanine.

Asparagine residue 19 appears to be a special case. As noted above, this residue is positioned directly across the cleft from lysine 111, and, while the latter residue is required for function, N19 can be replaced with either aspartic acid or alanine without disrupting BFP biogenesis. Given the crucial position of these residues with respect to the two domains of N-BfpC, we tested the hypothesis that either or both of these residues interacts with another component of the machine. Inspection of conserved domains of N-BfpE revealed several acidic residues that could potentially interact with K111. When we changed BfpE E83 to K, function was lost. Similarly when BfpC K111 was replaced with E, function was lost and when both changes were made simultaneously, function was not restored, so a simple specific interaction between these residues was not revealed. Mutation of BfpE D82 to N in

the setting of wild type BfpC similarly resulted in a loss of BFP biogenesis. However, in this case biogenesis was restored by the reciprocal modification of BfpC residue 19 from N to D. This extragenic suppression of a mutation in *bfpE* by the reciprocal mutation in *bfpC* implies that these residues interact directly. However, we could not further verify this prediction as N-BfpC and N-BfpE co-purified with BfpD in similar amounts regardless of whether the native or the N82D variant of N-BfpE was used. Furthermore this change in N-BfpE did not reduce BfpD activity.

From these studies it can be concluded that residue substitutions that profoundly affect the function of the pilus biogenesis machine may do so by mechanisms that do not preclude overall interactions among its components. Rather, it seems likely that such substitutions disrupt the orchestration of these interactions. It follows that interactions among components of T4P biogenesis machines must be precisely maintained for proper function. Herein we present the crystallographic structure of a T4bP machine component and reveal substantial similarity to T2S and T4aP components with which it shares negligible sequence similarity. The N-BfpC structure is a critical initial step in determining the architecture of the entire machine. An appreciation of this architecture is crucial to understanding the biogenesis of T4Ps both as an exemplary nanomachine and an important target for therapeutic intervention against diverse infectious diseases. Our evidence suggests that such machines may be inactivated without blocking interactions by interfering with specific residue functions.

Materials and Methods

Bacterial growth conditions and protein preparation

Bacterial strains (Supplementary Table 2) were routinely grown at 37 °C in Luria-Bertani broth with addition of ampicillin (200 µg/ml), chloramphenicol (20 µg/ml) and (kanamycin 100 µg/ml), as necessary to maintain plasmids. N-BfpC (amino acids 1–164) containing a C-terminal hexahistidine tag was purified to homogeneity from *E. coli* strains DH5αRPA302 as described.³⁵

Crystallization and structure determination

The Selenomethionine (Se-Met) substituted N-BfpC was expressed in *E. coli* B834 (DE3) in the presence of 50 mg/l of Se-Met and purified by the same procedure as native N-BfpC, except for using Sephacryl S-100 for the final size exclusion chromatography. Se-Met N-BfpC crystals were grown at room temperature by the hanging drop vapor-diffusion method using a reservoir solution containing 2.0 M ammonium sulfate. Crystals were in space group P2₁2₁2₁ with unit cell dimension a = 42.83 Å, b = 61.83 Å, c = 78.27 Å. Crystals were cryoprotected by transfer into the reservoir solution including 20% glycerol. MAD data were collected to a resolution of 2.1 Å at beamline 5.0.2 in Advanced light source. All data sets were processed with HKL2000 suite.⁴⁹

Phases were calculated with SOLVE⁵⁰ and the subsequent density modification using RESOLVE⁵¹ gave clear electron density. Models were manually built with XtalView⁵² with the partial model made by RESOLVE. Models were refined with simulated annealing, energy minimization, and individual B-factor refinement, using CNS.⁵³ The final *R* and *R*_{free} factors were 21.4 and 22.6 %, respectively. The model has good stereochemistry with 90.8 % in the most favored region, and no residues in the disallowed region, as assessed by PROCHECK.⁵⁴

Auto-aggregation assay

Plasmid pACYC184⁵⁵ was modified using the Stratagene QuikChange™ site-directed mutagenesis kit using primers Donne-1180 and Donne-1181 (Supplementary Table 3) to create plasmid pEM21 in which a *SacI* restriction site is inserted after position 1566. The *bfpC* gene was amplified from plasmid pMAR2⁵⁶ using primers Donne-1176 and Donne-1177 and cloned into the *SacI* and *XbaI* sites of pWKS130⁵⁷ from which it was subcloned into pEM21 to create plasmid pEM26. Variants of pEM26 and of plasmid pKDS302¹⁸, encoding the entire *bfp* operon, were made using QuikChange™. Overnight cultures of EPEC *bfpC* mutant strain UMD924²³ containing pEM26 or derivatives were diluted 50- to 100-fold in Dulbecco's modified Eagle's medium and grown for 4–6 h at 37°C before examination by phase contrast microscopy as described.³⁵ *E. coli cpxA** laboratory strain ALN924⁴² containing pKDS302 or its derivatives was diluted in Luria broth and grown for 4–6 h at 30°C with 0.1 mM IPTG before examination microscopically for auto-aggregation. A quantitative auto-aggregation assay was performed as previously described.³⁵

Flow cytometry

Overnight cultures of UMD924 with or without complementing plasmids were diluted 1:100 in DMEM/F12 without phenol red, and cultured at 37°C and 225 rpm for 4 h. Bacteria were pelleted, resuspended in PBS and fixed by heating at 65°C for 30 min. Alpha-1 bundlin antibody⁵⁸ was added to each sample at a 1:500 dilution in PBS and incubated at 4°C overnight. Anti-rabbit IgG, conjugated to Alexa Fluor 647 (Cell Signaling) secondary antibody at a concentration of 2 µg/ml was added to the bacteria and incubated at RT for 1 h. Cells were then stained with 1 µM SYTO 16 (Invitrogen) for 40 min at RT. Bacteria were analyzed for surface bundlin expression using a BD LSR II flow cytometer (BD Biosciences).

Localization of BfpD

Overnight cultures of *E. coli* strain TOP10 containing pLJC1, or strain DH5a containing pEM27, DH5a with pEM99, DH5a with pEM100, and DH5a with pEM101 were diluted (1:100) in 15 ml of LB with appropriate antibiotics and incubated at 37°C with shaking for 5 h. Cells were harvested by centrifugation for 20 min at ambient temperature and re-suspended in 3 ml of buffer containing 20 mM Tris-HCl pH 7.4, 140 mM NaCl, 1 mM phenylmethanesulfonyl fluoride (PMSF). Cells were passed twice through a French Press set to 20,000 pounds per square inch. Lysates were centrifuged at 12,000 ×g for 20 min. Recovered supernatants were centrifuged at 160,000 ×g for 40 min. Soluble fractions were saved. Tubes were rinsed twice with 1 ml of 20 mM Tris-HCl pH 7.4, 140 mM NaCl. Insoluble fractions were re-suspended in 3 ml of the same buffer and centrifuged at 160,000 ×g for 30 min. Supernatants were discarded. Tubes were rinsed twice with 1 ml of 20 mM Tris-HCl pH 7.4, 140 mM NaCl and re-suspended in 1 ml of the same buffer. Supernatants and insoluble fractions were stored at –80°C until analysis by immunoblotting with anti-GroEL antibody (Sigma) or the antibodies described below.

Co-affinity purification

E. coli strain DH5a cells transformed with relevant plasmids (Supplemental Table 2) were cultured in LB and stimulated with 1 mM IPTG at an Ab₆₀₀ of between 0.5 and 0.6. After 3 h, bacteria were pelleted and resuspended in lysis buffer containing 20 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole pH 7.4 and 1 mM PMSF. The cells were lysed in a French press (20,000 psi) and the insoluble proteins were removed from the lysate by centrifugation (10,000 ×g, 30 min, 4°C). The soluble fraction was incubated overnight at 4°C with nickel nitrilotriacetic acid (Ni-NTA) agarose beads and applied to a column. The flow-through was collected and the column was washed with 2 ml lysis buffer. This was followed with a 6 ml

wash step with lysis buffer containing 50 mM imidazole, and a 6 ml wash with lysis buffer containing 70 mM imidazole, the final 3 ml wash fraction was collected. The bound proteins were eluted with 1ml lysis buffer containing 250 mM imidazole, and collected. The collected samples from flow-through, wash and elute fractions were diluted with SDS-PAGE loading buffer, boiled for 5 min, separated by SDS/PAGE, transferred onto Immobilon-FL polyvinylidene membranes (Millipore) and incubated with Odyssey Blocking buffer (LI-COR Biosciences) overnight at 4°C. Membranes were probed with polyclonal rabbit anti-BfpC₁₋₁₆₄³⁵ at a dilution of 1:2000 or a sheep polyclonal antibody raised against the N-terminal amino acids 1–114 of BfpE at a dilution of 1:15000. Secondary goat anti-rabbit IRDye 800 (LI-COR) or rabbit anti-sheep IRDye 800 (Rockland) antibodies were used at dilutions of 1:20000. Proteins were imaged and band intensities were quantified by the Odyssey Infrared Imaging System (LI-COR). For determination of co-elution values, the band intensity of N-BfpC or N-BfpE in the elution fraction of each sample was expressed as a proportion of the respective total intensities from the flow-through, wash and elution fractions. Values from six assays were compared with those of the negative control, pEM170, without BfpD, and the positive control, pEM171, containing wild type N-BfpC, N-BfpE and BfpD.

ATPase activity

Samples containing wild-type or mutated versions of N-BfpC and N-BfpE with or without BfpD separated by co-affinity purification as described above, were assayed for ATPase activity using malachite green.⁵⁹ Protein samples were concentrated in an Amicon Ultra 3kDa column and quantified by spectrophotometry at Ab₂₆₀. Fifteen microliters of each quantified sample were added to 35 μl of assay buffer (0.02% (v/v) Triton X-100, 100mM Tris-HCl pH 7.4, 20mM KCl, 6mM MgCl₂ and 1mM ATP) and incubated at 37°C for 30min. The amount of free phosphate released upon ATP hydrolysis was measured by absorbance at 655nm following the addition of 10 μl of 4.2% (w/v) ammonium molybdate in 4 M sulfuric acid and 10 μl of 0.045% (w/v) aqueous malachite green and extrapolated from a standard curve generated using K₂HPO₄.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the staff of ALS beamline 5.0.1 for diffraction facilities. This work was supported by National Institutes of Health awards AI22160 (J.A.T.) and AI37606 (M.S.D.). AY was supported in part by a Skaggs Institute of Chemical Biology fellowship.

References

1. Craig L, Pique ME, Tainer JA. Type IV pilus structure and bacterial pathogenicity. *Nat. Rev. Microbiol.* 2004; 2:363–378. [PubMed: 15100690]
2. Ng SY, Wu J, Nair DB, Logan SM, Robotham A, Tessier L, Kelly JF, Uchida K, Aizawa S, Jarrell KF. Genetic and mass spectrometry analyses of the unusual type IV-like pili of the archaeon *Methanococcus maripaludis*. *J. Bacteriol.* 2011; 193:804–814. [PubMed: 21075925]
3. Collyn F, Lety MA, Nair S, Escuyer V, Ben Younes A, Simonet M, Marceau M. *Yersinia pseudotuberculosis* harbors a type IV pilus gene cluster that contributes to pathogenicity. *Infect. Immun.* 2002; 70:6196–6205. [PubMed: 12379698]
4. Essex-Lopresti AE, Boddey JA, Thomas R, Smith MP, Hartley MG, Atkins T, Brown NF, Tsang CH, Peak IR, Hill J, Beacham IR, Titball RW. A type IV pilin, PilA, contributes to adherence of *Burkholderia pseudomallei* and virulence in vivo. *Infect. Immun.* 2005; 73:1260–1264. [PubMed: 15664977]

5. Farinha MA, Conway BD, Glasier LMG, Ellert NW, Irvin RT, Sherburne R, Paranchych W. Alteration of the pilin adhesin of *Pseudomonas aeruginosa* PAO results in normal pilus biogenesis but a loss of adherence to human pneumocyte cells and decreased virulence in mice. *Infect. Immun.* 1994; 62:4118–4123. [PubMed: 7927665]
6. Tang H, Kays M, Prince A. Role of *Pseudomonas aeruginosa* pili in acute pulmonary infection. *Infect. Immun.* 1995; 63:1278–1285. [PubMed: 7890385]
7. Bieber D, Ramer SW, Wu CY, Murray WJ, Tobe T, Fernandez R, Schoolnik GK. Type IV pili, transient bacterial aggregates, and virulence of enteropathogenic *Escherichia coli*. *Science.* 1998; 280:2114–2118. [PubMed: 9641917]
8. Herrington DA, Hall RH, Losonsky G, Mekalanos JJ, Taylor RK, Levine MM. Toxin, toxin-coregulated pili, and the *toxR* regulon are essential for *Vibrio cholerae* pathogenesis in humans. *J. Exp. Med.* 1988; 168:1487–1492. [PubMed: 2902187]
9. Tacket CO, Taylor RK, Losonsky G, Lim Y, Nataro JP, Kaper JB, Levine MM. Investigation of the roles of toxin-coregulated pili and mannose-sensitive hemagglutinin pili in the pathogenesis of *Vibrio cholerae* O139 infection. *Infect. Immun.* 1998; 66:692–695. [PubMed: 9453628]
10. LaPointe CF, Taylor RK. The type 4 prepilin peptidases comprise a novel family of aspartic acid proteases. *J. Biol. Chem.* 2000; 275:1502–1510. [PubMed: 10625704]
11. Strom MS, Nunn DN, Lory S. A single bifunctional enzyme, PilD, catalyzes cleavage and N-methylation of proteins belonging to the type IV pilin family. *Proc. Natl. Acad. Sci. USA.* 1993; 90:2404–2408. [PubMed: 8096341]
12. Zhang H-Z, Donnenberg MS. DsbA is required for stability of the type IV pilin of enteropathogenic *Escherichia coli*. *Mol. Microbiol.* 1996; 21:787–797. [PubMed: 8878041]
13. 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]
14. 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; 378:32–38. [PubMed: 7477282]
15. 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* toxincoregulated pilus and *Pseudomonas aeruginosa* PAK pilin. *Mol. Cell.* 2003; 11:1139–1150. [PubMed: 12769840]
16. Hwang J, Bieber D, Ramer SW, Wu CY, Schoolnik GK. Structural and topographical studies of the type IV bundle-forming pilus assembly complex of enteropathogenic *Escherichia coli*. *J. Bacteriol.* 2003; 185:6695–6701. [PubMed: 14594844]
17. Tomich M, Planet PJ, Figurski DH. The *tad* locus: postcards from the widespread colonization island. *Nat. Rev. Microbiol.* 2007; 5:363–375. [PubMed: 17435791]
18. Stone KD, Zhang H-Z, Carlson LK, Donnenberg MS. A cluster of fourteen genes from enteropathogenic *Escherichia coli* is sufficient for biogenesis of a type IV pilus. *Mol. Microbiol.* 1996; 20:325–337. [PubMed: 8733231]
19. Donnenberg MS, Girón JA, Nataro JP, Kaper JB. A plasmid-encoded type IV fimbrial gene of enteropathogenic *Escherichia coli* associated with localized adherence. *Mol. Microbiol.* 1992; 6:3427–3437. [PubMed: 1362446]
20. Zhang H-Z, Lory S, Donnenberg MS. A plasmid-encoded prepilin peptidase gene from enteropathogenic *Escherichia coli*. *J. Bacteriol.* 1994; 176:6885–6891. [PubMed: 7961448]
21. Anantha RP, Stone KD, Donnenberg MS. The role of BfpF, a member of the PilT family of putative nucleotide-binding proteins, in type IV pilus biogenesis and in interactions between enteropathogenic *Escherichia coli* and host cells. *Infect. Immun.* 1998; 66:122–131. [PubMed: 9423848]
22. Ramer SW, Schoolnik GK, Wu CY, Hwang J, Schmidt SA, Bieber D. The type IV pilus assembly complex: biogenic interactions among the bundle-forming pilus proteins of enteropathogenic *Escherichia coli*. *J. Bacteriol.* 2002; 184:3457–3465. [PubMed: 12057939]
23. Anantha RP, Stone KD, Donnenberg MS. Effects of *bfp* mutations on biogenesis of functional enteropathogenic *Escherichia coli* type IV pili. *J. Bacteriol.* 2000; 182:2498–2506. [PubMed: 10762251]

24. Milgotina EI, Lieberman JA, Donnenberg MS. Corrigendum - The inner membrane subassembly of the enteropathogenic *Escherichia coli* bundle-forming pilus machine. *Mol. Microbiol.* 2011; 81:1125–1127. [PubMed: 21977501]
25. Planet PJ, Kachlany SC, DeSalle R, Figurski DH. Phylogeny of genes for secretion NTPases: identification of the widespread tadA subfamily and development of a diagnostic key for gene classification. *Proc. Natl. Acad. Sci. U. S. A.* 2001; 98:2503–2508. [PubMed: 11226268]
26. 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]
27. Blank TE, Donnenberg MS. Novel topology of BfpE, a cytoplasmic membrane protein required for type IV fimbrial biogenesis in enteropathogenic *Escherichia coli*. *J. Bacteriol.* 2001; 183:4435–4450. [PubMed: 11443077]
28. Abendroth J, Mitchell DD, Korotkov KV, Johnson TL, Kreger A, Sandkvist M, Hol WG. The three-dimensional structure of the cytoplasmic domains of EpsF from the type 2 secretion system of *Vibrio cholerae*. *J. Struct. Biol.* 2009; 166:303–315. [PubMed: 19324092]
29. Sauvonnet N, Gounon P, Pugsley AP. PpdD type IV pilin of *Escherichia coli* K-12 can be assembled into pili in *Pseudomonas aeruginosa*. *J. Bacteriol.* 2000; 182:848–854. [PubMed: 10633126]
30. Sauvonnet 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]
31. Kim TJ, Bose N, Taylor RK. Secretion of a soluble colonization factor by the TCP type 4 pilus biogenesis pathway in *Vibrio cholerae*. *Mol. Microbiol.* 2003; 49:81–92. [PubMed: 12823812]
32. Han X, Kennan RM, Parker D, Davies JK, Rood JI. Type IV fimbrial biogenesis is required for protease secretion and natural transformation in *Dichelobacter nodosus*. *J. Bacteriol.* 2007; 189:5022–5033. [PubMed: 17513472]
33. 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]
34. Sandkvist M, Bagdasarian M, Howard SP, DiRita VJ. Interaction between the autokinase EpsE and EpsL in the cytoplasmic membrane is required for extracellular secretion in *Vibrio cholerae*. *EMBO J.* 1995; 14:1664–1673. [PubMed: 7737119]
35. Crowther LJ, Anantha RP, Donnenberg MS. The inner membrane subassembly of the enteropathogenic *Escherichia coli* bundle-forming pilus machine. *Mol. Microbiol.* 2004; 52:67–79. [PubMed: 15049811]
36. Holm L, Sander C. Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.* 1998; 26:316–319. [PubMed: 9399863]
37. 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]
38. Karuppiiah V, Derrick JP. Structure of the PilM-PilN Inner Membrane Type IV Pilus Biogenesis Complex from *Thermus thermophilus*. *J. Biol. Chem.* 2011; 286:24434–24442. [PubMed: 21596754]
39. Yamagata A, Tainer JA. Hexameric structures of the archaeal secretion ATPase GspE and implications for a universal secretion mechanism. *EMBO J.* 2007; 26:878–890. [PubMed: 17255937]
40. Yeo HJ, Savvides SN, Herr AB, Lanka E, Waksman G. Crystal structure of the hexameric traffic ATPase of the *Helicobacter pylori* type IV secretion system. *Mol. Cell.* 2000; 6:1461–1472. [PubMed: 11163218]
41. Camberg JL, Johnson TL, Patrick M, Abendroth J, Hol WG, Sandkvist M. Synergistic stimulation of EpsE ATP hydrolysis by EpsL and acidic phospholipids. *EMBO J.* 2007; 26:19–27. [PubMed: 17159897]

42. Nevesinjac AZ, Raivio TL. The Cpx envelope stress response affects expression of the type IV bundle-forming pili of enteropathogenic *Escherichia coli*. J. Bacteriol. 2005; 187:672–686. [PubMed: 15629938]
43. Sandkvist M, Keith JM, Bagdasarian M, Howard SP. Two regions of EpsL involved in species-specific protein-protein interactions with EpsE and EpsM of the general secretion pathway in *Vibrio cholerae*. J. Bacteriol. 2000; 182:742–748. [PubMed: 10633109]
44. Ayers M, Sampaleanu LM, Tammam S, Koo J, Harvey H, Howell PL, Burrows LL. PilM/N/O/P proteins form an inner membrane complex that affects the stability of the *Pseudomonas aeruginosa* type IV pilus secretin. J. Mol. Biol. 2009; 394:128–142. [PubMed: 19857645]
45. Abendroth J, Kreger AC, Hol WG. The dimer formed by the periplasmic domain of EpsL from the Type 2 Secretion System of *Vibrio parahaemolyticus*. J. Struct. Biol. 2009; 168:313–322. [PubMed: 19646531]
46. 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]
47. Sampaleanu LM, Bonanno JB, Ayers M, Koo J, Tammam S, Burley SK, Almo SC, Burrows LL, Howell PL. Periplasmic domains of *Pseudomonas aeruginosa* PilN and PilO form a stable heterodimeric complex. J. Mol. Biol. 2009; 394:143–159. [PubMed: 19857646]
48. Pelicic V. Type IV pili: *e pluribus unum?* Mol. Microbiol. 2008; 68:827–837. [PubMed: 18399938]
49. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997; 276:307–326.
50. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr. D. Biol. Crystallogr. 1999; 55:849–861. [PubMed: 10089316]
51. Terwilliger TC. Reciprocal-space solvent flattening. Acta Crystallogr. D. Biol. Crystallogr. 1999; 55:1863–1871. [PubMed: 10531484]
52. McRee DE. XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. J. Struct. Biol. 1999; 125:156–165. [PubMed: 10222271]
53. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. D. Biol. Crystallogr. 1998; 54:905–921. [PubMed: 9757107]
54. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 1993; 26:283–291.
55. Chang ACY, Cohen SN. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. 1978; 134:1141–1156. [PubMed: 149110]
56. Baldini MM, Kaper JB, Levine MM, Candy DC, Moon HW. Plasmid-mediated adhesion in enteropathogenic *Escherichia coli*. J. Pediatr. Gastroenterol. Nutr. 1983; 2:534–538. [PubMed: 6352891]
57. Wang RF, Kushner SR. Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*. Gene. 1991; 100:195–199. [PubMed: 2055470]
58. Fernandes PJ, Guo Q, Donnenberg MS. Functional consequences of sequence variation in bundlin, the enteropathogenic *Escherichia coli* type IV pilin protein. Infect. Immun. 2007; 75:4687–4696. [PubMed: 17635860]
59. Lanzetta PA, Alvarez LJ, Reinach PS, Candia OA. An improved assay for nanomole amounts of inorganic phosphate. Anal. Biochem. 1979; 100:95–97. [PubMed: 161695]
60. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U. S. A. 2001; 98:10037–10041. [PubMed: 11517324]

- The structure of essential type IVb pilus component N-BfpC is presented.
- N-BfpC resembles type 2 secretion component EpsL more than type IVb component PilM.
- Non-functional BfpC variants retain recruitment, binding and activity of ATPase BfpD.

Figure 1.

N-BfpC fold, structure, and electrostatic implications for IM interactions. **(A)** N-BfpC fold. Domain I (blue) and II (green) show similar topologies with dissimilar regions (pink). The right view is rotated 90° from the left view. The central cleft is formed between $\alpha 1$ and $\alpha 2$. **(B)** N-BfpC electrostatic surface potential viewed as in Figure 1a (left) and rotated ~180° (right). Electrostatic potential was calculated using APBS,⁶⁰ where red represents negative charge, blue is positive and white is uncharged (scale -8 kT to $+8 \text{ kT}$). The positively charged patch near the C-terminal end (right view) and negatively charged regions are apparent. **(C)** Proposed configuration of N-BfpC with respect to IM membrane phospholipids. The transmembrane segment and the beginning of the periplasmic domain of BfpC (light blue), fold, and overlaid electrostatic surface potential as calculated with APBS,⁶⁰ is shown using PyMOL (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC).

Figure 2.

N-BfpC resembles N-EpsL and PilM. (A) Alignment of the amino acid sequences of EPEC N-BfpC and *V. cholerae* N-EpsL, adjusted to correspond to common secondary structure elements as aligned by the DALI server. Identical residues are shown in red. Domains I of N-BfpC and II of N-EpsL are colored in blue, domains II and III, respectively are colored in green, while cyan indicates the α -helical turn between the two domains. Cylinders correspond to α -helices and arrows to β -sheets, which in N-EpsL are numbered or lettered according to the original publication.³⁷ The second EpsL α -helix (asterisk) was resolved only when N-EpsL was co-crystallized with N-1 domain of EpsE.³³ Although the secondary structural features of PilM also align to those of BfpC, residues conserved between PilM and EpsL are not shared with BfpC. (B) Structural comparison of N-BfpC, N-EpsL (from PDB ID: 2BH1) and PilM (from PDB ID 2YCH) colored as in A. Domains I and II of N-BfpC correspond to domains II and III of N-EpsL and PilM (C) N-EpsL:N1-EpsE complex. The structure is rotated relative to its orientation in (B), with the top tilted toward the reader and the bottom tilted into the page. N1-EpsE (orange) is bound at the top of the cleft between domain II (light blue) and domain III (light green) of N-EpsL. Regions on N-EpsL that interact with N1-EpsE are colored magenta. (D) N-BfpC structure, rotated as for (C). Residues that correspond to those in regions 1 and 3 of N-EpsL are shown in stick representation, with carbon atoms colored magenta, oxygens in red and nitrogens in blue. The C-terminal peptide, which occupies a position in N-BfpC that corresponds to the PilN peptide in the PilM:PilN complex, is shown in stick representation with its carbons in yellow. (E) PilM:PilN complex rotated as for (C) showing the bound ATP. The PilN peptide, which binds in a cleft between domains I (grey) and II (light blue), is shown in stick representation with its carbons in yellow. Structural figures were prepared with PyMOL.

Figure 3.

Residues of N-BfpC required for function. (A) Side chains for altered residues are shown in stick representation in N-BfpC structure. (B) Immunoblotting of BfpC derivatives expressed *in trans* from complementing plasmids in EPEC *bfpC* mutant strain UMD924. Loading volumes were adjusted according to Ab_{600} measurements. Wild-type BfpC was expressed from plasmid pEM26; mutant proteins BfpC_{N19D}, BfpC_{N19A}, BfpC_{L21A}, BfpC_{N19A,L21A}, BfpC_{T28A}, BfpC_{K37A}, BfpC_{M42A}, and BfpC_{K111E} were expressed from plasmids pEM92, pKMS01, pEM93, pEM94, pEM95, pEM96, pEM97, and pEM98, respectively. Wild type EPEC strain E2348/60 was used as a positive control for native BfpC. The BfpC antibody, raised against the N-terminus of the protein, recognizes a BfpC degradation fragment as well as full-length BfpC. The auto-aggregation (AA) ability of each strain is indicated. (C) Auto-aggregation indexes, expressed as the percentage increase in OD_{600} after vortexing, of *bfpC* mutant UMD924 without (None, gray bar) or with complementing plasmids pEM26, pKMS01, pEM95, pEM96, or pEM92 encoding functional BfpC variants as indicated (black bars). Each assay was performed in triplicate in three experiments and error bars indicate standard errors of the means. Symbols: *, P value < 0.05, ***, P value < 0.001 in comparison to the uncomplemented mutant; †, P value < 0.05 in comparison to all other strains; §, P < 0.05 in comparison to all other functional variants. (D) Mean fluorescence intensity of strains capable (black bars) or incapable of auto-aggregation (gray bars) examined by flow-cytometry using an anti-bundlin antibody. E2348/69 is the wild type EPEC strain, UMD924 is the *bfpC* mutant and the remainder of the strains represent UMD924 containing plasmids encoding wild type *bfpC* (WT, pEM26) or plasmids pEM26, pEM92, pKMS101, pEM93, pEM94, pEM95, pEM96, pEM97 and pEM98 encoding BfpC variants with the respective indicated alterations. Values are the means of three separate experiments and error bars indicate standard errors of the means.

Figure 4.

Modifications of BfpC that interfere with function do not interfere with BfpD and BfpE interactions or BfpD activity. **(A)** Soluble (S) and insoluble (I) fractions of *E. coli* strain TOP 10 expressing only BfpD from pLJC1; or strain DH5 α expressing BfpC, BfpD and BfpE from pEM27; BfpC and BfpD from pEM122; BfpC_{L21A} and BfpD from pEM124, BfpC_{M42A} and BfpD from pEM128; and BfpC_{K111A} and BfpD from pEM129 were examined by immunoblot using anti-BfpD sera. **(B)** Soluble (S) and insoluble (I) fractions of *E. coli* strain DH5 α expressing BfpD only from pLJC1; BfpC, BfpD and BfpE from pEM27; BfpC_{L21A}, BfpD and BfpE from pEM99; BfpC_{M42A}, BfpD and BfpE from pEM100; and BfpC_{K111A}, BfpD and BfpE from pEM101 were examined by immunoblot using anti-BfpD sera. **(C)** The proportion of total N-BfpC (white bars, top), N-BfpE (gray bars, middle), and BfpD-His (black bars, bottom) in eluate fractions after elution from an Ni-NTA column of the indicated recombinant protein variants. Indicated plasmids encode WT N-BfpC and WT N-BfpE without BfpD-His (–, pEM170) or BfpD-His (+) and WT N-BfpC and WT N-BfpE (pEM171); N-BfpC_{M42A} and WT N-BfpE (pEM177); WT N-BfpC and N-BfpE_{D82N} (pEM179); N-BfpC_{N19D} and N-BfpE_{D82N} (pEM181); or N-BfpC_{N19A L21A} and WT N-BfpE (pEM184). Because the results were not normally distributed, N-BfpC and N-BfpE are shown as “box and whisker” plots where boxes indicate the interquartile range, bars within boxes indicate the medians and the “whiskers” indicate the range of six separate experiments. The asterisks indicate a significant difference compared to pEM171 ($P = 0.03$, Wilcoxon Matched-Pairs Signed-Ranks Test). For BfpD-His, the error bars indicate standard errors of the mean. **(D)** ATPase activity, in nanomoles of inorganic phosphate produced per minute per mg of total protein, of samples collected as in C. Samples collected from the strain containing pEM170 (no BfpD-His) were excluded because of extremely low protein concentrations. Error bars indicate the SEM of three separate experiments. **(E)** Phase-contrast micrographs of *E. coli* strain ALN92 harboring plasmids from which only bundlin and BfpG are expressed (–, pKDS301), all Bfp proteins are expressed (+, pKDS302) or all Bfp proteins are expressed with the following

modifications: BfpC_{N19D} (pEM38), BfpE_{D82N} (pEM39) and both BfpC_{N19D} and BfpE_{D82N} (pEM42). All images were taken at the same magnification.

Table 1

Crystallographic data and refinement statistics

Data collection	SeMetNBfpC		
	Peak	Edge	Remote
Beamline	ALS 5.0.1		
Space group	$P2_12_12_1$		
Cell dimensions	42.83, 61.83, 78.27		
α, β, γ (°)	90.0, 90.0, 90.0		
Resolution (Å)	20.0–1.90	20.0–2.0	20.0–2.10
Wavelength (Å)	0.9791	0.9793	0.9639
Completeness (%)	99.9 (99.6)	98.5 (99.4)	95.2 (94.0)
Observed reflections	67,381	56,740	49,574
Unique reflections	16,999	14,243	12,283
R_{sym} (%)^{a,b}	6.1 (24.7)	5.7 (23.6)	6.6 (25.8)
$I/\sigma(I)$	28.7 (4.3)	27.3 (5.6)	20.0 (3.3)

Refinement Statistics	
Resolution limits (Å)	20.0–1.9
Molecules/A.U.	1
R_{cryst} (%)^c	21.7
R_{free} (%)^d	22.7
No. of reflections used	16509
No. of atoms	
Protein	1266
Water	119
B-factor (Å²)	
Average	26.1
Protein	25.3
Water	34.6
RMS deviations	
Bond lengths (Å)	0.005
Bond angles (°)	1.24
Ramachandran plot	
Most favoured	92.9
Allowed	6.4
Generously allowed	0.7
Disallowed	0.0

^aValues in parentheses correspond to the highest resolution shell^b R_{sym} is the unweighted R value on I between symmetry mates: $\sum hkl \sum j |I(hkl) - \langle I(hkl) \rangle| / \sum hkl \sum I(hkl)$.^c $R_{\text{cryst}} = \frac{\sum hkl \|F_{\text{obs}}(hkl)\| - \|F_{\text{calc}}(hkl)\|}{\sum hkl \|F_{\text{obs}}(hkl)\|}$.

d_{Rfree} is the cross validation R factor for 5% of reflections against which the model was not refined.

Table 2

Summary of the effects of *bfpC* and *bfpE* mutations on function. Autoaggregation was assessed by the ability to complement an EPEC *bfpC* mutant strain in the case of pEM26 and derivatives or the ability to perform autoaggregation in the case of *E. colicpxA** strain ALN92 containing the entire *bfp* operon with or without mutations. For additional details see text.

BfpC substitutions	BfpE substitution	Autoaggregation	BfpD membrane recruitment*	Co-elution with BfpD
None	None	+	+	+
N19A	None	+	NT	NT
N19D	None	+	NT	NT
N19A, L21A	None	–	NT	+
L21A	None	–	+	NT
T28A	None	+	NT	NT
K37A	None	+	NT	NT
M42A	None	–	+	+
K111E	None	–	+	NT
None	D82N	–	NT	+
None	E83K	–	NT	NT
N19D	D82N	+	NT	+
K111E	E83K	–	NT	NT

* NT, not tested