Structure of the Bundle-forming Pilus from Enteropathogenic *Escherichia coli* 

Received for publication, July 25, 2005, and in revised form, September 15, 2005. Published, JBC Papers in Press, September 19, 2005, DOI 10.1074/jbc.M508099200

Stéphanie Ramboarina1,3, Paula J. Fernandes5, Sarah Daniell3, Suhail Islam2, Pete Simpson1,5, Gad Frankel4, Frank Booy5, Michael S. Donnenberg1, and Stephen Matthews1,2

From the 1Department of Biological Sciences, Wolfson Laboratory and the 6Center for Structural Biology, Imperial College London SW7 2AZ, United Kingdom and the 4Division of Infectious Diseases, School of Medicine, University of Maryland, Baltimore, Maryland 21201

Bundle-forming pili (BFP) are essential for the full virulence of enteropathogenic *Escherichia coli* (EPEC) because they are required for localized adherence to epithelial cells and auto-aggregation. We report the high resolution structure of bundlin, the monomer of BFP, solved by NMR. The structure reveals a new variation in the topology of type IVb pilins with significant differences in the composition and relative orientation of elements of secondary structure. In addition, the structural parameters of native BFP filaments were determined by electron microscopy after negative staining. The solution structure of bundlin was assembled according to these helical parameters to provide a plausible atomic resolution model for the BFP filament. We show that EPEC and *Vibrio cholerae* type IVb pilins display distinct differences in their monomer subunits consistent with data showing that bundlin and TcpA cannot complement each other, but assemble into filaments with similar helical organization.

A wide variety of pathogenic bacteria express fimbriae or pili, surface appendages that in many cases allow the microbe to attach to host epithelia. This attachment is often the first step toward colonization of their preferred host niche. The most important classes of pili are the chaperone-usher family and the type IV fimbriae. Type IV pilis are produced by a vast array of important human pathogens including *Vibrio cholerae, Neisseria gonorrhoeae, Neisseria meningitidis, Pseudomonas aeruginosa, Legionella pneumophila, Salmonella enterica serovar Typhi,* and enteropathogenic *Escherichia coli* (EPEC).3 A number of important phenotypes are associated with the expression of type IV pilis including auto-aggregation (1–3), adherence (4–6), twitching motility (7, 8), biofilm formation (9), horizontal gene transmission (10–12), cellular invasion (13, 14), and virulence (1, 15, 16).

The bundle-forming pili (BFP) of EPEC is an excellent model for the study of type IV pilis. BFP are required for the characteristic localized adherence of EPEC to epithelial cells (5, 17), for auto-aggregation and for full virulence in volunteers (1). The 14-gene bfp operon contains all of the genes specifically required for BFP biogenesis (18). Mutagenesis studies have shown that all but one of these genes is necessary to produce functional BFP (19–21).

The first gene of the bfp operon, bfpA, encodes prebundlin, which upon cleavage yields a 19-kDa protein that is the only known structural component of BFP (5). Prebundlin is associated with the inner membrane with its 12-amino acid N-terminal leader sequence in the cytoplasmic and C-terminal globular head domain located in the periplasm. Prebundlin is processed into mature bundlin after removal of the N-terminal leader sequence by the prepilin peptidase, BfpP (22). It has also been shown that DsbA, a periplasmic oxidoreductase stabilizes the globular head domain by catalyzing the formation of a well conserved disulfide bond in the C-terminal part of bundlin (23). Single substitutions of the two cysteine residues to serine are detrimental for pilus biogenesis and EPEC cell adhesion (23). The conserved disulfide bond is a common feature of type IV pilis (24). In addition, type IV pilis share sequence similarity in the fifty N-terminal amino acids, which have been assumed to form an extended, hydrophobic N-terminal helix that promotes the assembly of type IV fimbriae (25).

As BFP from EPEC represents an attractive system to study the assembly of type IV pilis, we have investigated the structure of the major subunit of BFP filaments, bundlin, by NMR. Here we report the high resolution structure of the 18.3-kDa globular head domain from bundlin and reveal a new variation in the topology of type IVb pilins. In combination with EM data collected on native BFP we were able to propose an atomic resolution model for the type IVb filament structure.

**EXPERIMENTAL PROCEDURES**

**NMR Spectroscopy**—The 18.3-kDa 15N-13C-labeled recombinant bundlin domain was expressed and purified as described previously (26). NMR samples of bundlin (350 μl) were prepared at a concentration of about 0.5 mM into 20 mM sodium phosphate buffer pH 5.2, 0.1% NaN3, and 3-(trimethylsilyl) propionic acid for referencing 1H, 15N, and 13C chemical shifts. NMR data were acquired at 303 K on a 500 MHz four-channel Bruker DRX500 spectrometer equipped with a z-shielded gradient triple resonance cryoprobe. Sequence-specific backbone and side chain assignments for bundlin were determined previously (26). Three-dimensional 500 MHz 1H-15N and 800 MHz 1H-13C NOESY-HSQC data were collected with a mixing time of 100 ms. Spectra were processed with NMRPipe (27) and analyzed using NMRview (28). 15N relaxation rate measurements were performed as previously described (29, 30) on a 800 MHz Varian Inova spectrometer. For the measurement of T1, ten experiments were performed with relaxation delays of 8.6, 9.7, 193, 996, 1493, 2000, 193, and 996 ms. Ten experiments were also performed for the measurement of T2 with relaxation delays of 8.5, 17.1, 34.1, 59.7, 76.8, 93.8, 119.4, 145.0, 17.1, and 76.8 ms. For T1 and T2, two experiments were repeated with identical delays to estimate the noise level. The heteronuclear 1H-15N NOE were meas-

---

1. This work was supported by The Wellcome Trust. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

2. *To whom correspondence may be addressed. E-mail: mdonnenb@umaryland.edu.*

3. *The abbreviations used are: EPEC, enteropathogenic *E. coli*; BFP, bundle-forming pilis; EM, electron microscopy; FT, fourier transformation; TMV, tobacco mosaic virus; PDB, Protein Data Bank.*
ured from two experiments with and without proton saturation. All the 
$^1$H-$^1$H HSQC spectra were processed with NMRview. Rates and errors 
were fitted as implemented in the NMRview5 software version.

**Structure Calculations**—Secondary structure elements in bundlin 
were first identified using the chemical shift-based, dihedral angle pre-
diction software TALOS (31). A manual analysis of $^15$N and $^13$C 
NOESY-HSQC spectra allowed the backbone NOEs characteristic of 
secondary structures to be identified. Additional long distance NOEs 
were assigned between side chain protons of residues Ile$^{37}$ and Tyr$^{156}$. 
The manually assigned NOEs were used to start the automated, itera-
tive, assignment procedure ARIA 1.2 (32, 33), which employed the 
structure calculation program CNS (34). Hydrogen bonds were mea-
sured from H$_2$O/D$_2$O exchange experiments and extracted from 
$^1$H-$^1$N HSQC spectra recorded every hour over 24 h. The disulfide bridge 
between Cys$^{116}$ and Cys$^{166}$ (numbering referenced to bundlin) was 
included in the calculations. 100 structures were imposed in the final 
itration, and 15 lowest energy structures with no distance and dihedral 
age violations greater than 0.2 Å and $5^\circ$, respectively were retained and 
water-refined in ARIA. A summary of NMR-derived restraints and sta-
tistics on the ensemble of structures is shown in TABLE ONE.

*FIGURE 1. The structure of the periplasmic domain of the bundlin from EPEC. A, plots of the 
$^1$N T1/T2 ratio (above) and $^1$H-$^1$N NOEs (below) 
versus the residue numbers of bundlin. Higher val-
ues of T1/T2 ratio of 29.34 and 40.12 up to 24.00 
obtained for residues Asp$^{170}$ and Ser$^{171}$ are not 
shown on the diagram. No data are shown for res-
ides Lys$^{167}$-Thr$^{168}$ as no signal was observed in 
the NMR spectra. The plot was generated using 
Jplot (C. Jeremy Craven, University of Sheffield); B, 
steroview of the backbone of bundlin for the 
elogue of 15 best structures superimposed 
after water refinement. C, ribbon representation 
of the secondary structural elements in bundlin. For 
clarity, only the $\alpha$-helices are annotated and col-
ored in blue, while the $\beta$-sheet is represented in 
cyan. D, ribbon representation of bundlin with the 
novel $\beta$1-strand indicated with an arrow.*
Structure of the Bundle-forming Pilus

**Bacterial Culture**—50 µl of EPEC bacterial culture, wild-type strain E2348/69, were used to seed growth at a 1:100 dilution in Hepes-modified Dulbecco’s modified essential medium and grown for 3 h, without shaking at 37 °C. Bacteria were collected by centrifugation at 4000 × g for 5 min at room temperature, and the pellet was resuspended in 50 µl of phosphate-buffered saline. The bacterial suspension was diluted 2-fold in phosphate-buffered saline, and TMV was added at 10 µg ml⁻¹ as a calibration control.

Bacteria were negatively stained using a modified version of Valentine’s technique. 2 µl of the bacterial/TMV suspension were pipetted between a thin carbon film and its mica support (1 mm²). The carbon film was washed by partially floating onto water in a teflon block and then floated off completely on 1% uranyl acetate stain. The carbon film was then collected on a lacey grid (Ted Pella) from below and viewed in an FEI CM200 TEM operating at 120 kV. Images of bacteria expressing BFP were recorded in low-dose mode (using Kodak SO163 film) at a magnification of ×36,000 with the first zero of the contrast transfer function in the range of 18–22 Å.

**Image Analysis**—Suitable micrographs were selected based on electron optical parameters (such as appropriate defocus, stigmatism, and absence of drift) and were digitized using a Nikon microtome at 6.35 µm/pixel. All computational analysis was carried out using IMAGIC software (35).

**Two-dimensional Projection Image**—Individual filaments were divided into segments of 350 Å in length, which were then normalized, band-pass filtered, masked, aligned, and summed to obtain a higher contrast, average two-dimensional projection image. In addition, larger areas containing multiple, naturally aligned BFP were selected, aligned, and summed, as above.

**Fourier Transformation**—The FTs of the individual pilus, the summed projection image and individual and summed areas of BFP were calculated and compared. All transforms resulted in common layer lines. The distance between the layer lines was measured to determine the axial repeat. A lattice was drawn on the diffraction pattern resulting from the summed FTs of individual BFP, which incorporated the major reflections. Bessel orders were calculated from the spacing of the reflections from the meridional line, as described under “Results.”

**Computational Assembly of Filament**—A space-filling fiber was built based on the NMR structure of the bundlin monomer and the helical parameters and measurements derived from the EM data. First, the 22-amino acid N-terminal helix was restored to the bundlin monomer. The orientation of the bundlin monomer was then carefully optimized by analogy with the type IV pilus models (36). The long N-terminal α-helix of the bundlin monomer was first oriented roughly parallel to the central axis of the pilus with a slight orientation of about 10° of the monomer with respect to the central axis to avoid clashes in the BFP filament. Furthermore, the bundlin monomer was oriented such that the N-terminal helix was positioned inside the helical strand, forming the innermost layer of the helical fiber. Hlxbuild software (situs.biomachina.org) was then used to assemble a single helix six subunits long. The helical rise was set to 22 Å; the radial displacement of each chain from the helical axis (x-offset in hlxbuild) was optimized to 23 Å, which gave an outer diameter close to that measured by EM (~75 Å) and avoided subunit clashes in the final structure.

The resulting helical strand was then duplicated by opening two further copies using Swiss-Pdb (37) to form a three-stranded helical fiber. Each copy was then translated by 44 and 88 Å, respectively. The helical strands were then merged to create the final representation.

**Coordinates Deposition**—Coordinates for the ensemble of NMR structures have been deposited at the Protein Data Bank under the accession code 1zwt. Tables of NMR assignments and restraints have been deposited in the BioMagResBank in Madison, WI under the accession code 6003.

**RESULTS**

**Description of the Structure of Bundlin**—Recombinant, double-labeled ¹³N-¹⁵C-mature bundlin lacking the hydrophobic N-terminal 22 residues was expressed and purified as described previously (26). Backbone resonance and side chain assignments were obtained using standard triple resonance NMR experiments (38). A correlation time of about 11.5 ns was calculated from ¹³N relaxation data (T1/T2 ratios along the sequence shown in Fig. 1A), which is consistent with that of the 18.3 kDa bundlin monomer. Overall ¹³N relaxation data for the backbone amide of bundlin reveal a well ordered structure with very few highly mobile

---

**TABLE ONE**

| Structural statistics on the 50 best NMR structures (PDB: 1zwt) |
|---------------------------------------------------------------|
|                                 |                      |
| Structural restraints (from Aria1.2*) | 577 |
| Ambiguous NOEs⁴                                      | 2277 |
| Unambiguous NOEs⁵                                    | 38 |
| Hydrogen bonds                                       | 87 |
| φ and ψ dihedral angles⁶                             | |
| Deviations from idealized geometry                   | |
| Bonds (Å)                                             | 0.0004 ± 0.0001 |
| Angles (degrees)                                      | 0.55 ± 0.03 |
| Improper (degrees)                                   | 1.73 ± 0.10 |
| Root mean square deviations (Å)                       | 1.09 ± 0.19 |
| Backbone (all residues)                               | 1.49 ± 0.18 |
| Heavy atoms (all residues)                            | 1.79 ± 0.14 |
| Backbone on secondary structure                       | 1.04 ± 0.14 |
| Heavy atoms on secondary structure                    | |
| Ramachandran plot (from Procheck NMR)                 | |
| Most favored regions                                  | 77% |
| Additionally allowed regions                          | 18% |
| Generously allowed regions                            | 3% |
| Disallowed regions                                    | 2% |

⁴ Derived using Aria1.2, and no distance violation greater than 0.5 Å occurred in the final ensemble.

⁵ Derived from TALOS, and no violation greater than 5° occurred in the final ensemble.
regions (Fig. 1). Interestingly, resonance assignments were not achieved for residues Lys$^{167}$, Asn$^{168}$, and Thr$^{169}$ as signals were not observed in the NMR spectra (numbering referenced to bundlin). Furthermore, high T1/T2 fluctuations are observed for the region between residues 162 and 173 without a decrease of $^1$H-$^1$H NOE (Fig. 1A). This is indicative of conformational exchange for the protein backbone in this region as is also evidenced by $^1$H-$^1$H NOE cross-peaks of much lower intensity than those exhibited by the ordered parts of the protein.

The ensemble of 15 low energy structures with neither distance nor dihedral violations more than 0.2 Å or 5°, respectively is shown in Fig. 1B. The three-dimensional structure of the globular head domain of bundlin is composed of a seven-stranded mixed parallel and antiparallel $\beta$-sheet surrounded by four $\alpha$-helices and two $3_1\alpha$-helices (Fig. 1C). The backbone root mean square deviation is 0.79 ± 0.14 Å for the secondary structure elements and 1.09 ± 0.19 Å for all atoms (TABLE ONE). The four helices encompass residues Lys$^{32}$–Tyr$^{51}$ (α1-C), Leu$^{60}$–Asn$^{67}$ (α2),

FIGURE 2. Electron microscopy of BFP from EPEC. A, electron micrograph of negatively stained wild-type EPEC expressing BFP. B, naturally occurring, aligned bundles of BFP shown at a higher magnification. Thicker filaments are TMV whose 23-Å spacing was used as a calibration standard. C, projection of a two-dimensional average of aligned, summed, short individual segments of BFP. The pitch of the main 3-start helix is indicated, together with the approximate diameter of the pilus. Protein is shown as white. Scale bars: 200 nm.

FIGURE 3. EM analysis. A, FT resulting from the sum of individual FTs of straight sections of BFP. The meridional and equatorial lines are labeled. The layer line peaks used to calculate the Bessel orders and their associated pitch are indicated by arrows. B, central region of the FT shown in A with a two-dimensional lattice indicating the major reflections.
Structure of the Bundle-forming Pilus

Lys\textsuperscript{113}–Ala\textsuperscript{120} (α3), and Thr\textsuperscript{158}–Ala\textsuperscript{165} (α4) with α3 and α4 conjoined by the disulfide bond between Cys\textsuperscript{116}–Cys\textsuperscript{166}. The loop region (52–59) connecting α1 and α2 displays a higher degree of flexibility in bundle structure than the rest of the molecule as shown by a decrease of \( ^{1}H-N\) NOEs within this region (Fig. 1A). In contrast, the second loop (68–80) connecting α2 to the β-sheet possesses a short 3\(_{10}\)-helix from Asp\textsuperscript{75}–Tyr\textsuperscript{77} that is rigidly associated with the rest of the secondary structure.

The β-strands are defined by residues Lys\textsuperscript{41}–Thr\textsuperscript{46} (β1), Glu\textsuperscript{90}–Ala\textsuperscript{96} (β2), Tyr\textsuperscript{104}–Thr\textsuperscript{109} (β3), Lys\textsuperscript{129}–Val\textsuperscript{135} (β4), Ser\textsuperscript{141}–Glu\textsuperscript{145} (β5), Ala\textsuperscript{151}–Ser\textsuperscript{154} (β6), and Asn\textsuperscript{172}–Met\textsuperscript{179} (β7). The first three β-strands do not form the β-sheet that is usually observed in type IV pilins (see below) (25, 39–41). Instead they are arranged in a complex mixed β-sheet topology, with the following strand order β2/β3/β1 (Fig. 1D).

This arrangement is unusual and results principally from hydrophobic interactions between Ile\textsuperscript{61} from β1, and Pro\textsuperscript{86}, Leu\textsuperscript{108}, and Leu\textsuperscript{111}, which are located in the β1/β2-loop, β3 and β3/α3-loop, respectively. Helix α3 connects β3 to β4, which forms an antiparallel pairing with neighboring strands, β6 and β7. β4 is followed by a short 3\(_{10}\)-helix (Glu\textsuperscript{138}–Asn\textsuperscript{140}) and the β5-strand, which lies antiparallel to β6. In summary, all seven strands form a twisted contiguous surface arranged in the order β2, β3, β1, β7, β4, β6, and β5 (Fig. 1C and D).

The β-sheet region of bundlein is flanked by the four α-helices on three sides. The α1-C helix is positioned across the concave face of the sheet by hydrophobic interactions between Ile\textsuperscript{67} and Tyr\textsuperscript{104}, Thr\textsuperscript{44} and Tyr\textsuperscript{104}, and Ala\textsuperscript{49}–Phe\textsuperscript{67}. Residues Ile\textsuperscript{67}, Tyr\textsuperscript{51}, Tyr\textsuperscript{57}, Leu\textsuperscript{60}, Ile\textsuperscript{64}, Val\textsuperscript{93}, and Pro\textsuperscript{95} delineate a hydrophilic platform located between the first two helices and the first part of the β-sheet. Significant hydrophobic interactions involving residues Ile\textsuperscript{67}, Tyr\textsuperscript{51}, Tyr\textsuperscript{57}, and Tyr\textsuperscript{104} lead to a well-defined orientation for α2 of 90° relative to α1-C. The long loop, 68–80, following α2 also adopts a fixed orientation (Fig. 1C and D), mediated by hydrophobic interactions between residues Leu\textsuperscript{85}, Pro\textsuperscript{72}, Tyr\textsuperscript{75}, and Tyr\textsuperscript{105}.

Two important hydrophobic cores appear to stabilize the protein. The first hydrophobic core comprising residues within β1, β3/α3-loop, α3, β4, α4, and β7, is delineated by Ile\textsuperscript{65}, Leu\textsuperscript{111}, Ala\textsuperscript{114}, Leu\textsuperscript{119}, Tyr\textsuperscript{131}, Val\textsuperscript{143}, Ala\textsuperscript{165}, and Tyr\textsuperscript{177}. Another cluster of hydrophobic residues, located on the opposite face, encompasses residues Phe\textsuperscript{102}, Tyr\textsuperscript{105}, Ile\textsuperscript{135}, Ile\textsuperscript{141}, Phe\textsuperscript{144}, Ala\textsuperscript{152}, and Phe\textsuperscript{178}. As a result the well ordered β4/β5-loop is found close to the β2/β3-loop in the bundlein structure (Fig. 1C and D). Residues forming the hydrophobic cores are highly conserved among the different EPEC strains.

Analysis of BFP by Electron Microscopy—BFP expressed by wild-type EPEC were viewed after negative staining (Fig. 2, A and B). The majority of BFP was expressed in early stage growth and formed closely associated swathes of pili emanating from the bacterial membrane. Clearly discernable individual pili were chosen for further analysis. Short sections corresponding to ~350 Å in length were selected, band-pass filtered, normalized, aligned, and summed, resulting in the two-dimensional-averaged structure shown in Fig. 2C. The image clearly shows a repeating diagonal pattern of protein densities (white) along the pili, which is indicative of a helical structure. From this averaged projection image, the pili are estimated to have a diameter of about 75 Å. From this structure, the pitch of the dominant helix was estimated to be 44 Å, as shown in Fig. 2C. Fourier transforms were calculated from the two-dimensional projection image, from individual pili and also from larger areas selected from the naturally occurring aligned bundles of pili. All Fourier transforms have major layer lines in common.

The FT shown in Figs. 3, A and B is derived from the sum of transforms calculated from individual pili and was used to deduce the helical parameters of BFP. All pili were centered prior to calculation of their FT.

This summed pattern was used as the layer lines are most clearly represented. A two-dimensional lattice was drawn on the FT to intercept the major reflections (Fig. 3B). The distances between these reflections and the equatorial line were then measured, as shown. Three main reflections are visible with spacings of 1/88 Å, 1/44 Å, and 1/22 Å on layer lines 4, 8, and 12, respectively. The distance between layer lines was measured as 1/175 Å, which corresponds to the axial repeat defined by the predicted 1-start helix. The 1/22 Å meridional spacing was only visible in some FTs, but defines the axial rise of the subunits as 22 Å. The structure of the Type IVb pilin proteins are characterized by three regions, the αβ-loop, core and the D-region.

![Figure 4. Comparison of the topology between the structures of the type IVb pilin proteins](Image 314x353 to 564x732)
Bundlin Structure Provides New Insights into the Type IV Family—
The NMR structure of bundlin, the major subunit of BFP from EPEC, adopts an αββ fold composed of four α-helices that surround a platform formed by a seven-stranded, twisted, contiguous β-sandwich. Our study reveals a new topology for the mixed parallel and antiparallel β-strands arranged in the following order β2, β3, β1, β7, β4, β6, and β5. Although common characteristics can be observed between bundlin and other type IV pilin structures, the new structural features reported here for bundlin have not been described for other members of the type IV pilin family.

The type IV pilin family is divided into two groups, types IVa and IVb, distinguished by the length of the pre-pilin leader sequence and the size of the pilin protein (36). The first high resolution structures of pilin proteins were those from the type IVa class, GC from N. gonorrhoeae (42), PAK and K122-4 from P. aeruginosa (39, 40). The type IVa pilin structure is characterized by an extended N-terminal α-helix and a globular head domain that is folded into an α-β-roll configuration (39, 40, 42). The globular head domain is well conserved and characterized by a four-stranded antiparallel continuous β-meander among all the type IVa pilins (36). The structures of TcpA from V. cholerae and PilS from S. typhi from the type IVb group have been determined recently by x-ray diffraction (25) and NMR, respectively (41). The type IVb pilin proteins are characterized by a longer leader peptide, larger globular head domain and display a significantly different topology than the type IVa proteins (25, 36). The globular head domain of bundlin also exhibits a novel variation of the arrangement of β-sheet and α-helices. Furthermore, crystallographic analysis of the pseudopilin PulG from Klebsiella oxytoca reveals the four internal anti-parallel β-strands characteristic of type IV pilins, but the highly variable loop region with a disulfide bond is not found (43).

Type IVb pilins are characterized by two regions, the αβ-loop and the D-region. The αβ-loop includes the first two N-terminal α-helices (α1 and α2), whereas the D-region is delineated by helices α3 and α4, which are linked by the conserved disulfide bond (25, 36). Comparison of the bundlin structure with that of TcpA and PilS (Fig. 4) reveals some structural similarities in the arrangement of strands in the central region of the β-sheet, but bundlin adopts several unique structural features. The structure-based sequence alignment and superimposition of bundlin, TcpA, and PilS are shown in Figs. 5 and 6.

Both similarities and striking differences are observed among the topologies of the three proteins. The αβ-loop region of the three proteins differs in size and in the distribution of secondary structure elements. The αβ-loop in TcpA is defined by the two N-terminal α-helices, α1 and α2, followed by a long loop (25, 36) whereas an additional antiparallel β-strand is observed in the equivalent loop of PilS (41) (Figs. 5 and 6). The shorter αβ-loop region in bundlin, composed of α1-C and α2, is significantly more structured and displays an altered orientation with respect to the long N-terminal helix. In contrast to the αβ-loop, fragments in the D-region show some structural equivalence within the three pilins, i.e. within α3, α4, and the short region encompassing the β4- and β6-strands in bundlin (Figs. 5 and 6). It is noteworthy that, after the superimposition of the equivalent regions between bundlin and both TcpA and PilS, some hydrophobic positions in the D region appear to be conserved between the three proteins (Fig. 6).

A common architectural characteristic of TcpA and PilS is the inclusion of the C-terminal β-strand in the middle of the structure forming an antiparallel arrangement. In contrast, the C-terminal β-strand β7 in bundlin, forms a mixed parallel and anti-parallel β-sheet with the adjacent β1- and β4-strands, because of the novel insertion of the short β1-strand between β3 and β7 (Fig. 4). As expected, the structure-based sequence alignment reveals that the fragment containing the β1-strand in bundlin is absent in both TcpA and PilS sequences (Figs. 5 and 6). Despite this major difference, equivalence can be established between β2 of bundlin and β1 and β3 of TcpA and PilS, respectively and among the C-terminal β-strands of all three pilins. The region delineated by β3 of bundlin, β2 in TcpA and β4 in PilS appears to be equivalent in the three proteins with some conserved hydrophobic residues (Fig. 6).

BFP Model and Biological Implications—By combining the monomer structure with the native structure from electron microscopy we are able to present a model of the pilus. This provides some insight into the function of this important virulence factor. Our NMR data reveal that exclusion of the first thirty amino acids of bundlin (“the N-terminal hydrophobic helix”) impairs the ability of the pilin to assemble into BFP. Helical reconstruction of BFP was constrained by experimental data obtained from EM of negatively stained filaments and, as suggested in
other models of Type IV pili (36, 43), the positioning of the N-terminal hydrophobic helix within the central cavity of the pilus. Although the flexibility or bending of the N-terminal helix cannot be assessed from our data, our proposed model of BFP filament displays an innermost layer formed by a coiled-coil structure of the highly conserved N-terminal helices (Fig. 7B) (36). Subunit-subunit interactions between globular head domains involve residues from the αβ-loop and D-region and are located on the edges of the outermost layer in the fragment immediately preceding α2-, and the α4-helix, respectively (Fig. 7C). The 3-start left-handed helix of our model is similar to other recently published models and suggests that type IV pili could be assembled from three sites simultaneously. This concept is consistent with the recent observation that alternating subunits of the hexameric BfpD ATPase engage the pilus biogenesis machine, implying there are three sites at which pilin subunits are added concurrently (44). Although the helical parameters determined by EM for TCP and BFP filaments are very similar, local interactions and forces that maintain the structural integrity of the pilus are likely to be different as we observe striking differences within the αβ-loop and D-regions. For instance, the α2-helix within the αβ-loop from the BFP is more solvent-exposed and not involved in significant subunit-subunit contacts. This is consistent with observations that replacing key elements of bundlin with the corresponding regions from TcpA renders EPEC unable to produce pili (45). An analysis of sequence variations in bfpA revealed nine different alleles that can be divided into two major classes termed α and β (46, 47). Sequence variations were concentrated in the C terminus. In the BFP structure, the majority of these variable residues are also located in the outer rim of the pilus (Fig. 7E). In particular a region of the bfpA gene encoding amino acids 115–152 has an excess of non-synonymous substitutions, indicating that this portion of the protein is under selective pressure for diversification. One potential source of this selective pressure is the host immune response. Consistent with this hypothesis, this

FIGURE 6. Structure-based sequence alignment between bundlin, TcpA, and PilS. Sequence alignment was based on the results given by Dali between bundlin and TcpA. Alignment between bundlin and PilS was also reported as structural equivalence can be observed. Schemes of secondary structure corresponding to each three-dimensional structure are represented above each primary sequence. For the sake of clarity, only the α-helices and β-strands were annotated in bundlin. Residues highlighted in blue are identical in the three proteins. Residues shown in red correspond to hydrophobic residues found in the same position within the three proteins. Conserved cysteines are labeled in green.
region is surface-exposed in our BFP model (Fig. 7E). Antibodies that recognize epitopes within this region might be protective against EPEC infection and thereby select for EPEC variants that have altered epitopes.

One of the more striking findings in this study is the marked difference in secondary and tertiary structure among pilin proteins from the type IVb pilus family. Despite these differences the pilin genes have recognizable sequence similarities and the pili themselves have virtually identical structures and function. Evidently, powerful evolutionary forces contribute to diversification in pilin structure while pilus structure and function are maintained.

Acknowledgments—We thank Geoff Kelly and Tom Frenkiel of the 800 MHz NMR service at NIMR.

REFERENCES
1. Bieber, D., Ramer, S. W., Wu, C. Y., Murray, W. J., Tobe, T., Fernandez, R., and Schoolnik, G. K. (1998) Science 280, 2114–2118
2. Kirn, T. J., Lafferty, M. J., Sandoe, C. M., and Taylor, R. K. (2000) Mol. Microbiol. 35, 896–910
3. Park, H. S., Wolfgang, M., van Putten, J. P., Dorward, D., Hayes, S. F., and Kooney, M. (2001) Mol. Microbiol. 42, 293–307
4. Doig, P., Saxtry, P. A., Hodges, R. S., Lee, K. K., Paranchych, W., and Irvin, R. T. (1990) Infect. Immun. 58, 124–130
5. Donnenberg, M. S., Giron, J. A., Nataro, J. P., and Kaper, J. B. (1992) Mol. Microbiol. 6, 3427–3437
6. Stone, B. J., and Abu Kwaik, Y. (1998) Infect. Immun. 66, 1768–1775
7. Merz, A. J., So, M., and Sheetz, M. P. (2000) Nature 407, 98–102
8. Bradley, D. E. (1980) Can. J. Microbiol. 26, 146–154
9. O’Toole, G. A., and Kolter, R. (1998) Mol. Microbiol. 30, 295–304
10. Seifert, H. S., Ajioka, R. S., Marchal, C., Sparling, P. F., and So, M. (1988) Nature 336, 392–395
11. Yoshida, T., Furuya, N., Ishikura, M., Isobe, T., Haino-Fukushima, K., Ogawa, T., and Komano, T. (1998) J. Bacteriol. 180, 2842–2848
12. Chen, L., and Dubnau, D. (2003) Front. Biosci. 8, s544–s556
13. Donnenberg, M. S. (1996) Bacterial Invasiveness, Vol. 209, pp. 79–98
14. Zhang, X. L., Tsui, I. S., Fung, A. W., Wong, D. K., Dai, X., Yang, Y., Hackett, J., and Morris, C. (2000) Infect. Immun. 68, 3067–3073
15. Herrington, D. A., Hall, R. H., Losonsky, G., Mekalanos, J. J., Taylor, R. K., and Levine, M. M. (1988) J. Exp. Med. 168, 1487–1492
16. Farinha, M. A., Conway, B. D., Glasier, L. M. G., Ellett, N. W., Irving, R. T., Sherburne, R., and Paranchych, W. (1994) Infect. Immun. 62, 4118–4123
17. Giron, J. A., Ho, A. S. Y., and Schoolnik, G. K. (1991) Science 254, 710–713
18. Stone, K. D., Zhang, H. Z., Carlson, L. K., and Donnenberg, M. S. (1996) Mol. Microbiol.

FIGURE 7. Atomic resolution model for BFP. A, representation of the 3-start helical filament of BFP in which each subunit is represented as a sphere and the three strands in blue, yellow, and red. The 44-Å pitch, the 22-Å rise, and the diameter of about 85 Å are shown on the model. B, model of the BFP filament with each bundlin subunit shown as a ribbon representation. Helical strands are shown in yellow, blue, or red. C, surface representation of the model of BFP where the αp-loop, the core, and the D-regions are colored in green, blue, and yellow, respectively. The N-terminal helix forming the central layer of the filament is shown in gray. D, surface representation of the bundlin monomer showing the variable residues between α and β bundlin alleles. The orientation has been chosen such that the hydrophobic helix is at the back. Residues among bfpA alleles that show an excess of non-synonymous variations and appear to be under selective pressure for diversification are colored in red, whereas others variable regions are colored blue. E, surface representation of the model of BFP showing variable residues between α and β bundlin alleles. Residues among bfpA alleles that show an excess of non-synonymous variations and appear to be under selective pressure for diversification are colored in red, whereas other variable regions are colored blue.
Structure of the Bundle-forming Pilus

19. Anantha, T. P., Stone, K. D., and Donnenberg, M. S. (2000) J. Bacteriol. 182, 2498–2506
20. Hwang, J. W., Bieber, D., Ramer, S. W., Wu, C. Y., and Schoolnik, G. K. (2003) J. Bacteriol. 185, 6695–6701
21. Schreiber, W., Stone, K. D., Strong, M. A., DeTolla, L. J., Hoppert, M., and Donnenberg, M. S. (2002) Microbiology-Sgm 148, 2507–2518
22. Zhang, H. Z., Lory, S., and Donnenberg, M. S. (1994) J. Bacteriol. 176, 6885–6891
23. Zhang, H. Z., and Donnenberg, M. S. (1996) Mol. Microbiol. 21, 787–797
24. Strom, M. S., and Lory, S. (1993) Annu. Rev. Microbiol. 47, 565–596
25. Craig, L., Taylor, R. K., Pique, M. E., Adair, B. D., Arvai, A. S., Singh, M., Lloyd, S. J., Shin, D. S., Getzoff, E. D., Yeager, M., Forest, K. T., and Tainer, J. A. (2003) Mol. Cell 11, 1139–1150
26. Ramboarina, S., Fernandes, P., Simpson, P., Frankel, G., Donnenberg, M., and Matthews, S. (2004) J. Biomol. NMR 29, 427–428
27. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) J. Biomol. NMR 6, 277–293
28. Johnson, B. A., and Blevins, R. A. (1994) J. Biomol. NMR 4, 603–614
29. Kay, L. E., Torchia, D. A., and Bax, A. (1989) Biochemistry 28, 8972–8979
30. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Formanek, J. D., and Kay, L. E. (1994) Biochemistry 33, 5984–6003
31. Cornilescu, G., Delaglio, F., and Bax, A. (1999) J. Biomol. NMR 13, 289–302
32. Nilges, M., and O’Donoghue, S. I. (1998) Prog. NMR. Spectrosc. 32, 107–139
33. Linge, J. P., O’Donoghue, S. I., and Nilges, M. (2001) Methods in Enzymology: Nuclear Magnetic Resonance of Biological Macromolecules, Part B, Vol. 339, pp. 71–90, Academic Press, San Diego, CA
34. Brungar, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921
35. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R., and Schatz, M. (1996) J. Struct. Biol. 116, 17–24
36. Craig, L., Pique, M. E., and Tainer, J. A. (2004) Nat. Rev. Microbiol. 2, 363–378
37. Gregg, N., and Petitsch, M. C. (1997) Electrophoresis 18, 2714–2723
38. Sattler, M., Schleucher, J., and Griesinger, C. (1999) Prog. NMR Spectros. 34, 93–158
39. Hazes, B., Sastry, P. A., Hayakawa, K., Read, R. J., and Irvin, R. T. (2000) J. Mol. Biol. 299, 1005–1017
40. Keizer, D. W., Slupsky, C. M., Kalisiak, M., Campbell, A. P., Crump, M. P., Sastry, P. A., Hazes, B., Irvin, R. T., and Sykes, B. D. (2001) J. Biol. Chem. 276, 24186–24193
41. Xu, X. F., Tan, Y. W., Lam, L., Hackett, J., Zhang, M. J., and Mok, Y. K. (2004) J. Biol. Chem. 279, 31599–31605
42. Parge, H. E., Forest, K. T., Hickey, M. J., Christensen, D. A., Getzoff, E. D., and Tainer, J. A. (1995) Nature 378, 32–38
43. Kohler, R., Schäfer, K., Müller, S., Vignon, G., Diederichs, K., Philippens, A., Ringler, P., Pugsley, A. P., Engel, A., and Welte, W. (2004) Mol. Microbiol. 54, 647–664
44. Crowther, L. J., Yamagata, A., Craig, L., Tainer, J. A., and Donnenberg, M. S. (2005) J. Biol. Chem. in press
45. McNamara, B. P., and Donnenberg, M. S. (2000) Microbiology-Uk 146, 719–729
46. Blank, T. E., Zhong, H. L., Bell, A. L., Whittam, T. S., and Donnenberg, M. S. (2000) Infect. Immun. 68, 7028–7038
47. Blank, T. E., Lacher, D. W., Scaletsky, I. C. A., Zhong, H. L., Whittam, T. S., and Donnenberg, M. S. (2003) Emerg. Infect. Dis. 9, 113–115