Structural Characterization of CFA/III and Longus Type IVb Pili from Enterotoxigenic Escherichia coli

Subramaniapillai Kolappan, Justin Roos, Alex S. W. Yuen, Owen M. Pierce, and Lisa Craig

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

The type IV pili are helical filaments found on many Gram-negative pathogenic bacteria, with multiple diverse roles in pathogenesis, including microcolony formation, adhesion, and twitching motility. Many pathogenic enterotoxigenic Escherichia coli (ETEC) isolates express one of two type IV pili belonging to the type IVb subclass: CFA/III or Longus. Here we show a direct correlation between CFA/III expression and ETEC aggregation, suggesting that these pili, like the Vibrio cholerae toxin-coregulated pili (TCP), mediate microcolony formation. We report a 1.26-Å resolution crystal structure of CofA, the major pilin subunit from CFA/III. CofA is very similar in structure to V. cholerae TcpA but possesses a 10-amino-acid insertion that replaces part of the α2-helix with an irregular loop containing a 3_10-helix. Homology modeling suggests a very similar structure for the Longus LngA pilin. A model for the CFA/III pilus filament was generated using the TCP electron microscopy reconstruction as a template. The unique 3_10-helix insert fits perfectly within the gap between CofA globular domains. This insert, together with differences in surface-exposed residues, produces a filament that is smoother and more negatively charged than TCP. To explore the specificity of the type IV pili assembly apparatus, CofA was expressed heterologously in V. cholerae by replacing the tcpA gene with that of cofA within the tcp operon. Although CofA was synthesized and processed by V. cholerae, no CFA/III filaments were detected, suggesting that the components of the type IVb pili assembly system are highly specific to their pilin substrates.

Received 21 February 2012 Accepted 12 March 2012
Published ahead of print 23 March 2012
Address correspondence to Lisa Craig, lcraig@sfu.ca.
Supplemental material for this article may be found at http://jb.asm.org/.
Copyright © 2012, American Society for Microbiology. All Rights Reserved.
doi:10.1128/JB.00282-12
found almost exclusively on enteric bacterial pathogens, whereas type IVa pili are found on human pathogens with diverse tissue and organ specificities, including *Pseudomonas aeruginosa*, *Haemophilus influenzae*, and pathogenic *Neisseria* spp. Structural studies on three full-length type IV piliins (10, 11, 21, 40) and several recombinant piliins lacking the N-terminal polymerization domain (3, 10, 22, 31, 37, 45, 58) have revealed a common architecture: an extended 53-residue α-helix (α1), the C-terminal half of which is embedded in a globular domain with a central antiparallel β-sheet. Yet, the type IVa and IVb piliins have distinct protein folds. In the type IVa piliins, the β-sheet has continuous nearest-neighbor connectivity and is followed by a C-terminal loop that lies at the periphery of the globular domain, whereas the C terminus of the type IVb piliins is buried in the globular domain as the central strand of the β-sheet. The significance of the different protein folds is not known and may be a consequence of evolution within a common niche or of adaptation to a particular environment.

Pseudo-atomic resolution structures of two pilus filaments, the type IVa GC pilus from *Neisseria gonorrhoeae* and the type IVb TCP from *V. cholerae*, have been determined by transmission electron microscopy (EM) and image reconstruction followed by fitting of pilin crystal structures into the EM density envelope (11, 29). In both GC pili and TCP, pilin subunits are held together primarily by hydrophobic interactions among the N-terminal α-helices, which form a helical array in the core of the filament. The pilin globular domains are loosely packed on the filament surface, anchored by the N-terminal α-helices. Thus, the most conserved region of the protein, the N-terminal α-helix, mediates filament formation, whereas the more-variable globular domain is exposed and defines the diverse interactions and functions of the pili. The helical symmetries of the filaments are similar for GC pili and TCP, suggesting that the assembly mechanism is conserved between the two type IV pilus subtypes.

The type IV pilus biogenesis process is not well understood. The type IVa pili assembly requires as many as 40 gene products encoded in different parts of the genome, whereas the type IVb system comprises less than a dozen proteins, all encoded within the same approximate gene cluster. The type IVa and IVb piliins, which are homologous, suggest that the assembly mechanism is conserved between the two type IV pilus subtypes.

The type IV pilus biogenesis process is not well understood. The type IVa pili assembly requires as many as 40 gene products encoded in different parts of the genome, whereas the type IVb system comprises less than a dozen proteins, all encoded within the same approximate gene cluster. The type IVa and IVb piliins, which are homologous, suggest that the assembly mechanism is conserved between the two type IV pilus subtypes.

**Bacterial strains and plasmids.** Bacterial strains, plasmids, and primers are listed in Table 1. *V. cholerae* strains O395 and SK70 and ETEC strains 31-10 and 31-10p were provided by Ronald Taylor (Dartmouth Medical School). *E. coli* strains were grown in Luria-Bertani (LB) broth at 37°C. *V. cholerae* strains were grown under TCP-expressing conditions in LB broth (starting pH, 6.5) for 16 h at 30°C (26). ETEC strains were grown overnight in colonization factor antigen (CFA) agar (1% Casamino Acids [Difco], 0.15% yeast extract, 0.005% MgSO4, and 0.0005% MnCl2) with 2% agar added [16] at 37°C. Antibiotics were used at final concentrations of 100 μg/ml streptomycin, 45 μg/ml kanamycin, 30 μg/ml gentamicin, and 100 μg/ml ampicillin.

**Electron microscopy.** ETEC strains 31-10 and 31-10p were grown overnight on CFA agar plates. Cells were applied to glow-discharged carbon-coated copper grids (EM Sciences) by gently touching a single colony with the grid. Cells were stained with 1% phosphotungstic acid and imaged on a Hitachi 8000 STEM microscope at 5,000× magnification.

**Autoaggregation assay.** ETEC strains 31-10 and 31-10p were grown overnight on CFA agar plates for induction of CFA/III pili. Cells were scraped from the plate and gently suspended in phosphate-buffered saline (PBS, pH 6.0 and 7.0) or Tris-buffered saline (TBS, pH 8.5) with various concentrations of NaCl (10, 100, or 200 mM) and left to settle for 30 to 60 min. Ten-microliter aliquots were placed on glass slides, covered with a glass cover slip, and sealed with nail enamel. Cell aggregation was screened using light microscopy, and phase-contrast microscopy images were collected using a Leica DMI 4000B microscope at 10× magnification.

**Expression and purification of N-terminally truncated CofA.** The gene fragment encoding CoA residues 29 to 208 was PCR amplified from ETEC 31-10 genomic DNA using primers Ec-coa-f-primer and Ec-coa-r-primer and cloned into pET15b (Novagen) at the BamHI and NdeI sites to produce an N-terminal hexahistidine tag (His tag) for purification by metal affinity chromatography. N-terminally truncated CoA (ΔN-CoA) was expressed in *Escherichia coli* Origami(DE3) cells. Cells were grown to an optical density (at 600 nm) of 0.4 in LB broth containing ampicillin at 37°C. ΔN-CoA expression was induced with 0.4 mM isopropyl-β-D-1-thiogalactopyranoside, and cells were grown for a further 18 h at 19°C. Cells were harvested by centrifugation at 5,000 × g for 30 min. Cell pellets were flash-frozen in liquid nitrogen and stored at −80°C until required. Frozen cell pellets from 6-liter cultures were thawed and suspended in lysis buffer containing 50 mM NaH2PO4/Na2HPO4 (pH 7.4), 500 mM NaCl, and EDTA-free protease inhibitor cocktail (Roche). The supernatant was loaded onto a HisTrap column (GE Healthcare) preequilibrated with buffer A (50 mM NaH2PO4/Na2HPO4 [pH 7.4], 50 mM imidazole, 500 mM NaCl). The column was then washed with buffer A, and protein was eluted with buffer B (50 mM NaH2PO4/Na2HPO4 [pH 7.4], 500 mM NaCl, 50 to 300 mM imidazole in a linear gradient). Fractions containing the protein were pooled and dialyzed against buffer C (20 mM Tris-HCl [pH 7.4], 50 mM NaCl, 1 mM EDTA).
TABLE 1  Bacterial strains, plasmids, and primers

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Description or sequence</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacterial strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V. cholerae O395</td>
<td>O1 classical Ogawa, wild type, Sm&lt;sup&gt;+&lt;/sup&gt;</td>
<td>53</td>
</tr>
<tr>
<td>E. coli S17 7</td>
<td>piri pro recA hsdR17( r&lt;sub&gt;K&lt;/sub&gt; m&lt;sub&gt;K&lt;/sub&gt; ) [ RP4-2-Tc::Mu::Km&lt;sup&gt;+&lt;/sup&gt; Tn7 ] piri; Tp&lt;sup&gt;+&lt;/sup&gt; Sm&lt;sup&gt;+&lt;/sup&gt;</td>
<td>13</td>
</tr>
<tr>
<td>V. cholerae SIK70</td>
<td>O395; ΔtpA; Sm&lt;sup&gt;+&lt;/sup&gt;</td>
<td>R. K. Taylor</td>
</tr>
<tr>
<td>V. cholerae SJ003</td>
<td>SJ003, pMini1</td>
<td>This study</td>
</tr>
<tr>
<td>V. cholerae TR003</td>
<td>O395, ΔtpA coFA</td>
<td>This study</td>
</tr>
<tr>
<td>ETEC 31-10</td>
<td>Wild-type ETEC, carries 55-kb virulence plasmid with cof operon encoding CFA/III</td>
<td>24</td>
</tr>
<tr>
<td>ETEC 31-10p</td>
<td>ETEC 31-10, spontaneous CFA/III-deficient strain lacking 55-kb virulence plasmid</td>
<td>51</td>
</tr>
<tr>
<td>**E. coli Origami(DE3)</td>
<td>Δara-leu7697 ΔlacX74 ΔphoA PsvII phoA ard139 aplC galE galK rpsL F&lt;sup&gt;−&lt;/sup&gt; lac&lt;sup&gt;r&lt;/sup&gt; pro&lt;sup&gt;−&lt;/sup&gt; (DE3) gori522;Tn10 trxB; Km&lt;sup&gt;+&lt;/sup&gt; Sm&lt;sup&gt;+&lt;/sup&gt; Tc&lt;sup&gt;−&lt;/sup&gt;</td>
<td>Novagen</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pTK1</td>
<td>pKAS32::tcpA&lt;sup&gt;−&lt;/sup&gt;-rlL; Ap&lt;sup&gt;+&lt;/sup&gt;</td>
<td>26</td>
</tr>
<tr>
<td>pTK1-ΔBbsd</td>
<td>pTK1, BbsI site deleted</td>
<td>This study</td>
</tr>
<tr>
<td>pJR2</td>
<td>pTK1::coFA within tcpA&lt;sup&gt;−&lt;/sup&gt; genomic flanking sequences</td>
<td>This study</td>
</tr>
<tr>
<td>pMini1</td>
<td>pACYC184; Gm&lt;sup&gt;−&lt;/sup&gt; Mm&lt;sup&gt;−&lt;/sup&gt;</td>
<td>38</td>
</tr>
<tr>
<td>pMT5</td>
<td>toxA; Ap&lt;sup&gt;+&lt;/sup&gt;</td>
<td>14</td>
</tr>
<tr>
<td><strong>Primers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ec-coFA-fpcr</td>
<td>GGAATTCCATATGGACTCACTGACTGTTTCGTGATTTGTC</td>
<td>This study</td>
</tr>
<tr>
<td>Ec-coFA-rPCR</td>
<td>CCGGATCCCTTAACCGCTGCGCAAAATAGTAGGTTA</td>
<td>This study</td>
</tr>
<tr>
<td>oOMP-B2</td>
<td>CATGGAAGACCGGCTTCAAAATAGGTGTGTTAACTTAACTTTACG</td>
<td>This study</td>
</tr>
<tr>
<td>oOMP-R1</td>
<td>CATGGAAGACCGGCTTCAAAATAGGTGTGTTAACTTAACTTTACG</td>
<td>This study</td>
</tr>
<tr>
<td>oOMP-F1</td>
<td>CATGGAAGACCGGCTTCAAAATAGGTGTGTTAACTTAACTTTACG</td>
<td>This study</td>
</tr>
<tr>
<td>oOMP-R2</td>
<td>CATGGAAGACCGGCTTCAAAATAGGTGTGTTAACTTAACTTTACG</td>
<td>This study</td>
</tr>
<tr>
<td>Vc-tcpA-flank-f2</td>
<td>GCAATGTACATGACGATGATCGTG</td>
<td>This study</td>
</tr>
<tr>
<td>Vc-tcpA-flank-R2</td>
<td>GTTAAATGGTGGAGTATCAATAATAGC</td>
<td>This study</td>
</tr>
</tbody>
</table>

The His tag was removed by thrombin cleavage, and ΔN-CoFA was further purified on a Sephacryl S-100 HR size exclusion column (GE Healthcare) pre-equilibrated with buffer C. Fractions were analyzed by SDS-PAGE, and peak fractions were combined and concentrated to 15 mg/ml by using an Amicon stirred cell concentrator with a 3,500-Da molecular mass cutoff filter (Millipore). Protein was flash-frozen in liquid nitrogen and stored at −80°C.

**Crystallization and X-ray data collection.** Crystals of ΔN-CoFA were grown in hanging drops by vapor diffusion in 100 mM Tris-HCl (pH 8.5), 80°C. Protein was flash-frozen in liquid nitrogen and stored at −80°C.

**Structure determination and refinement.** The ΔN-CoFA structure was solved by molecular replacement using the N-terminally truncated V. cholerae TcpA crystal structure (PDB ID 1Q0Y) as the search model. Only the core β-sheet and α-helices of TcpA were used for the model. All non-identical residues between the CoFA and TcpA globular domains were changed to Ser, Ala, or Gly using COOT (15). PHASER (49) yielded two distinct solutions with relatively high Z-scores (8.61 and 7.10). Several cycles of simulated annealing refinement (Cartesian annealing) were performed using CNS (6), which provided an interpretable electron density map with two molecules in the unit cell. Further refinement using REFMAC5 (36) significantly improved the quality of the map, revealing electron density for most of the missing side chains and also the main chain of missing loops. The map was further improved by iterative rounds of fitting and refinement using COOT and REFMAC5 (36). The ARP/wARP program (42) was used to locate water oxygen atoms, which were confirmed by visual inspection of the electron density map and of a composite annealed omit map, calculated using CNS. Final cycles of restrained individual anisotropic refinement brought the R<sub>work</sub> and R<sub>free</sub> values to 0.160 and 0.194, respectively. The model was validated with PROCHECK, SFCHECK, and Molprobity (12, 27, 55) (Table 2).

**Construction of the CFA/III filament model.** The CFA/III pilus filament model was generated using the V. cholerae TCP structure, obtained by fitting of the TcpA crystal structure into our EM reconstruction volume (29). First, the N-terminal 28 residues that are missing in the CoFA crystal structure were modeled using the coordinates for the corresponding residues in the full-length PAK pilin crystal structure, as was done for TcpA. The full-length CoFA structure was then superimposed onto a TcpA subunit in the TCP model using the model-building program in COOT (15). The symmetry operators determined for the TCP reconstruction (8.4-Å axial rise, 96.8° azimuthal rotation) were applied to the subunit to generate a CFA/III filament model using COOT.

**Generation of V. cholerae cofA strain.** Seamless mutagenesis with the type IIS restriction enzyme BbsI (32) was used to replace the V. cholerae O395 tcpA gene fragment encoding the mature TcpA protein with that encoding mature CoFA in the suicide vector pTK1, a derivative of pKAS32, which carries the tcpA gene and its genomic flanking regions (48). First, pTK1 was mutagenized to delete a single BbsI site using QuikChange (Stratagene). Next, the resultant vector, pTK1-ΔBbsd, was linearized by inverse PCR amplification using primers oOMP-B2 and oOMP-R1 (Table 1). This step added BbsI digestion sites to each end of the PCR product (PCR fragment A) and eliminated the sequence encoding the mature TcpA protein, leaving the tcpA gene-flanking regions, including the segment encoding the TcpA signal peptide. The 626-bp coFA gene encoding the mature pilin protein was PCR amplified from ETEC 31-10 genomic DNA using primers oOMP-fA1 and oOMP-rA2, both of which carry BbsI digestion sites to each end of the PCR product. The PCR product was cloned into the Bluescript KS<sup>−</sup> vector and the E. coli Origami(DE3) strain used to express the recombinant protein. The resulting strain was used to generate a V. cholerae cofA strain.
TABLE 2 Crystallographic data collection and analysis

<table>
<thead>
<tr>
<th>Data collected</th>
<th>(\Delta N)-CofA value(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beamline</td>
<td>NSLS X26C</td>
</tr>
<tr>
<td>Space group</td>
<td>P1</td>
</tr>
<tr>
<td>Cell a, b, c (Å)</td>
<td>29.50, 34.15, 64.23</td>
</tr>
<tr>
<td>Cell α, β, γ</td>
<td>95.7, 97.4, 94.0</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50–1.26</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.82</td>
</tr>
<tr>
<td>Completeness (%)(a)</td>
<td>96.01/87.56</td>
</tr>
<tr>
<td>Wilson B-value (Å(^2))</td>
<td>10.9</td>
</tr>
<tr>
<td>No. of observed reflections</td>
<td>263,810</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>66,018</td>
</tr>
<tr>
<td>(R)(_{cryst}) (%)(b)</td>
<td>16.0</td>
</tr>
<tr>
<td>(R)(_{free}) (%)(c)</td>
<td>19.4</td>
</tr>
<tr>
<td>Avg B-factor (Å(^2))</td>
<td>14.4</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>2625</td>
</tr>
<tr>
<td>No. of water molecules</td>
<td>510</td>
</tr>
<tr>
<td>RMSD bond angles</td>
<td>0.009</td>
</tr>
<tr>
<td>RMSD bond lengths</td>
<td>1.336</td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
</tr>
<tr>
<td>Most favored (%)</td>
<td>90.6</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>9.4</td>
</tr>
</tbody>
</table>

\(a\) Overall/last shell (resolution range for last shell is 1.31–1.26).
\(b\) \(R\)\(_{cryst}\) is the weighted R value on I between symmetry mates.
\(c\) \(R\)\(_{free}\) is the cross validation R factor for 5% of reflections against which the model was not refined.

Production of anti-CofA antibodies. The \(\Delta N\)-CofA protein used for crystallization was used as an antigen for polyclonal rabbit anti-CofA antibody production (Pacific Immunology).

Figure preparation. Molecular structure figures were prepared using MacPyMol (http://www.pymol.org).

Protein structure accession number. The atomic coordinates for \(\Delta N\)-CofA have been deposited in the Protein Data Bank under accession number 3S0T.

RESULTS

ETEC autoaggregation correlates with CFA/III expression. To establish a role for CFA/III in ETEC autoaggregation, ETEC strains were grown overnight on CFA agar, which induces CFA/III expression (16). ETEC strain 31-10 cells, which carry the 55-kb virulence plasmid containing the cof operon, are highly piliated under these conditions, as shown by transmission electron microscopy (TEM) (Fig. 1A, left). In contrast, ETEC strain 31-10p cells, which lack this plasmid, displayed no pili on their surfaces when grown in the same medium (Fig. 1A, right). ETEC 31-10 cells grown under various other conditions, including CFA broth, LB agar, and LB broth, displayed very few pili (data not shown). Cells from CFA plates were suspended in PBS (pH 6.0 or 7.0) or TBS (pH 8.5) with various concentrations of NaCl (10 mM, 100 mM, or 200 mM) and allowed to settle, and then aggregation was observed using light microscopy or phase-contrast microscopy. ETEC 31-10 cells aggregated into large clusters under all buffer conditions tested (Fig. 1B, left). In contrast, no aggregation was observed for ETEC 31-10p grown on CFA plates (Fig. 1B, right) or for strain 31-10 grown in liquid culture. Thus, ETEC autoaggregation correlates well with CFA/III display.

Crystal structure of ETEC CofA. To characterize the CFA/III pilus, we determined the crystal structure of N-terminally truncated CofA (\(\Delta N\)-CofA, residues 29 to 208) to 1.26-Å resolution by molecular replacement with V. cholerae \(\Delta N\)-TcpA O1, classical bio-

Analysis of pilin and pilus expression in the V. cholerae cofA strain.

Expression of CofA was assessed by growing V. cholerae JR003 under pilus-inducing conditions (LB, pH 6.5, 30°C, 16 h on a Ferris wheel rotator), analyzing 20 μl of cell culture by SDS-PAGE, and immunoblotting with rabbit anti-CofA antibodies. For comparison, ETEC 31-10 cells expressing CofA and ETEC-31-10p, which lack the cof operon, were grown in CFA broth at 37°C shaking for 16 h. Pilius assembly was assessed by physically shearing the pili from the cells by using an Ultra-Turrax T8.01 disperser, filtering the cell homogenate, and analyzing it by SDS-PAGE and immunoblotting. DnaK was included as a loading control and blotted with anti-DnaK antibodies (Enzo Life Sciences). Pilius assembly was also assessed by EM imaging of a negatively stained V. cholerae JR003 cell culture. Five-microliter aliquots of overnight culture were applied to carbon-coated copper grids (Cedarlane Labs), excess fluid was wicked away after 2 min, and the grids were stained with 1% phosphotungstic acid, pH 7.0. Cells were imaged using an FEI F20 Supertwin electron microscope.

FIG 1 ETEC pilus expression correlates with bacterial aggregation. (A) TEM images of ETEC strains 31-10 and 31-10p after overnight growth on CFA agar plates. Inset: close-up of pili on 31-10 cells. Scale bar, 0.5 μm. (B) Phase-contrast images of ETEC 31-10 and 31-10p grown on CFA agar, suspended in PBS (pH 7.0, 100 mM NaCl), and allowed to settle for 60 min.
Although CofA and TcpA share 37% amino acid sequence identity overall, they are only 30% identical for the globular domains that were crystallized (Fig. 2A). Thus, a number of modified N-TcpA models were generated and used as search models to obtain a molecular replacement solution for N-CofA. N-CofA crystallized in the P1 space group with two molecules in the unit cell. All residues were readily fit into the electron density map for molecule A, whereas residues 113 to 115 are disordered in molecule B. As the two molecules are highly similar (root mean square deviation [RMSD], 0.39 Å for main chain atoms), only molecule A is shown.

CofA has the characteristic fold of the type IVb pilins with an N-terminal α-helix, α1C (residues 29 to 53), and a twisted 5-stranded antiparallel β-sheet with non-nearest-neighbor connectivity, where the C-terminal 8-residue segment forms the central strand (β5) (Fig. 2B). Between α1C and the β-sheet is the αβ-loop (residues 54 to 104), which begins with a broad loop (residues 54 to 68), followed by a 1.5-turn 3_10-helix (residues 69 to 73), a 2-turn α-helix (α2, residues 79 to 86), an extended irregular loop (87 to 105) containing a single-turn 3_10-helix (residues 92 to 94), and then a short β-hairpin-like loop with a type II turn. The αβ-loop is followed by the first two strands of the central β-sheet, β1 and β2, between which lies a large loop (residues 109 to 119). Following β2 the polypeptide chain exits the β-sheet to form a 4-turn α1-helix (α3, residues 129 to 142) that is distorted at its C-terminal end by two glycines, Gly137 and Gly140. α3 runs almost parallel to α1C and interacts with α1C and the β-sheet via hydrophobic side chains. After α3 the chain reverses direction to form β3. The 27-residue-long loop that follows β3 (the β3-β4 loop) winds around the “front” surface of the globular domain and buries much of the β-sheet. This loop feeds into the edge strand of the β-sheet (β4, residues 175 to 177), after which the polypeptide chain reverses direction via a type II
turn, followed by an extended stretch along the edge of the globular domain. The chain turns 90° to form a 1.5-turn α-helix, α4 (residues 189 to 194). Finally, an irregular loop connects α4 with β5, the central strand of the β-sheet (residues 201 to 208). The C-terminal residue, Arg208, forms a salt bridge with Asp144 at the N terminus of β3. Conserved cysteines at positions 132 and 196 form a disulfide bond that links the N terminus of α3 with the loop that follows α4. The D-region, defined as the 64-residue section between the two cysteines, and the β3-loop flank the β-sheet and together comprise a large portion of the globular domain. These two regions display considerable sequence and structural variability among the type IV pilins and house residues critical to pilus functions (8).

The N-terminal 28 residues of CofA (α1N) are not present in the ΔN-CofA structure but are predicted based on sequence homology to form the first half of an extended α-helix, α1, as seen in the full-length *P. aeruginosa* PAK pilin and *N. gonorrhoeae* GC pilin structures (10, 11). In the PAK and GC pilins, α1 is a continuous 53-residue α-helix with a curved S-shape. The C-terminal half of α1, α1C, is embedded in the globular domain, and the N-terminal half, α1N, protrudes like a stalk. The curvature of α1 in PAK and GC pilins is induced by three helix-disrupting residues, Gly14, Pro22, and Pro/Gly42. CofA also has a glycine at position 14, but it has isoleucine and threonine at positions 22 and 42, respectively. However, CofA has 3 glycine residues within positions 14, 22, and 42, respectively. Although the sequence similarities between CofA and TcpA are most notable difference between CofA and TcpA is a 10-amino-acid insertion in the αβ-loop of CofA that diverts the polypeptide chain between residues 59 and 78 relative to the TcpA structure. In TcpA, the loop emerging from α1 folds into the 4-turn α-helix, α2, which lies at a right angle to α1. In CofA, the polypeptide follows the same path as TcpA for residues 54 to 58, but this loop continues downward and outward relative to α1 and then loops back to form a 310-helix before entering a short α2 corresponding to the C-terminal half of α2 in TcpA. We refer to the region within the αβ-loop of CofA that diverges from TcpA as the 310 insert (residues 59 to 78). The two pilin structures also differ somewhat in the large irregular β3-β4 loop on the surface of the globular domain, which interacts with the 310 insert in CofA, and between β4 and the disulfide-bonded C196 (Fig. 2B and C). These regions, together with the 310 insert, are the least-well-conserved in sequence between the two proteins.

**CFA/III filament model.** Because of the structural similarities between CofA and TcpA, we used the TCP EM reconstruction to derive a model for the CFA/III pilus. In the TCP subunit, CofA subunits are arranged along a right-handed one-start helix with an axial rise of 8.4 Å and an azimuthal rotation of 96.8° per subunit (29). A full-length CofA subunit was first generated by appending the N-terminal 28 residues of the full-length PAK pilin structure (PDB ID 1OQW; 39% sequence identity, 75% sequence similarity over these residues) onto ΔN-CofA and changing individual residues *in silico* to the corresponding CofA residues. Full-length CofA was then superimposed onto a single TcpA subunit within the TCP filament, and the CFA/III filament was generated by imposing the helical symmetry operators on this subunit (Fig. 3A). The CofA subunits pack well in the CFA/III model, with no major steric clashes, as expected due to their structural similarity to TcpA. Importantly, the 310 insert fits into the cavity that lies between the globular domains, narrowing this gap considerably (Fig. 3A to D). The CofA αβ-loops are oriented inward and contact the N-terminal α-helix of their neighboring subunits, whereas the D-regions lie on the filament surface (Fig. 3D). The D-regions do not protrude from the filament surface to the extent that they do in TCP, and this feature, together with the 310 insert that partially fills the gaps between subunits, produces a smoother filament surface for CFA/III than for TCP. Because the N-terminal α-helix is not present in the CofA and TcpA crystal structures and this region is not well-resolved in the TCP EM reconstructions (29), its position in CFA/III is not known with precision. Nonetheless, this segment, as modeled using the corresponding PAK pilin coordinates, packs exceptionally well into the filament core, almost filling it completely without introducing major steric clashes (Fig. 3E).

The largest subunit-subunit interface occurs among the N-terminal α-helices in the hydrophobic core of the filament, but there are also a number of complementary charged residues on the globular domains that are in close proximity in neighboring subunits and are likely to form salt bridges. We had previously identified TcpA residues Arg26 in α1N and Glu83 in the αβ-loop as being important for filament assembly (30). These residues are located within salt bridging distance in the TCP model in neighboring subunits in the left-handed three-start helix. In the CFA/III model, Arg26 NH1 is within 5.2 Å of Glu93 Oε1, which corresponds to Glu83 in TcpA (Fig. 4). Small adjustments in dihedral angles of the side chains of these residues would allow formation of a stabilizing salt bridge in CFA/III. Salt bridges may also be present between charged residues within the CofA globular domains: Asp29 and Arg208 in the right-handed 4-start helix and Arg53 and Asp141 in neighboring subunits in the three-start helix (Fig. 4). In addition, the invariant Glu5 is located at the same axial position and within salt-bridging distance of the positively charged N-terminal amine in the adjacent subunit in the 1-start helix. These complementary charges occur at the same level in the filament due to the helical symmetry, with each subunit staggered along the 1-start helix by 8.4 Å. A salt bridge between these two residues would neutralize their charges in the otherwise-hydrophobic environment of the filament core (Fig. 4) and may provide a driving force for filament assembly (11).

**Computational model of LngA.** We used the ΔN-CofA crystal structure to derive a homology model for ΔN-LngA, the major pilin subunit of the second ETEC type IVb pilus, Longus (Fig. 5). Due to their high sequence identity (76%) LngA is predicted to be very close in structure to CofA and to possess the 310 insert within the αβ-loop. LngA has two fewer amino acids than CofA, which results in a shorter loop immediately preceding β4 (Fig. 5A and B, variable loop). Based on the amino acid similarities between CofA and LngA, the Longus filament is also expected to closely resemble that of CFA/III. Residues predicted to be involved in subunit-subunit interactions in CFA/III are well-conserved between the two pilins. Amino acid differences occur mainly on the solvent-exposed surface within the D-region. Nonetheless, both the shape and the chemistry of this exposed face are similar for CofA and LngA, and both have an overall negative charge (Fig. 5C). Amino
acid differences between the ETEC pilins and *V. cholerae* TcpA also occur mainly on the solvent-exposed surface. These differences, including those in the α9-loop that form the 310 insert in the ETEC pilins, result in different surface topographies and chemistries for TcpA and the ETEC pilins (Fig. 5C and D) and their corresponding pilus surfaces.

**Heterologous expression of CofA in V. cholerae.** Because of the structural similarity between CofA and TcpA and the sequence...
conservation in the interfaces between pilin subunits in the pilus filament models, we hypothesized that the \textit{V. cholerae} TCP bio-
genesis apparatus would assemble CFA/III pili if provided with the CoFA subunit. The \textit{V. cholerae} tcpA gene within the tcp operon was replaced with the \textit{cofA} gene by allelic exchange, leaving the seg-
ment encoding the TcpA signal peptide in place to allow for pro-
cessing by the \textit{V. cholerae} prepilin peptidase TcpJ. CofA expression in this \textit{V. cholerae} strain, JR003, was analyzed by SDS-PAGE and
immunoblotting, and CFA/III assembly was assessed by EM and
by shearing the pili off the cell surfaces. CofA is expressed in \textit{V. ch}
olerae JR003 and processed to its mature form when grown un-
der pilus-inducing conditions (Fig. 6A). However, the CofA band in
sheared cell supernatants was almost undetectable, whereas a
strong band was observed for ETEC 31-10 (Fig. 6B). No pilus
filaments were observed by EM for JR003, whether grown under
TCP-inducing conditions (LB, pH 6.5, 30°C) or CFA/III-inducing
conditions (CFA agar, 37°C) (data not shown). These results dem-
onstrate that CofA is synthesized and processed by \textit{V. cholerae},
but the TCP assembly apparatus cannot efficiently incorporate CoFA
into pilus filaments.

**DISCUSSION**

We have provided here evidence that CFA/III pil mediates ETEC
aggregation. ETEC cells aggregate when the 55-kb virulence plas-

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig5}
\caption{Comparison of the CoFA and LngA amino acid sequences and structures. (A) Amino acid sequence alignment of CoFA and LngA (ETEC strain E9034A; GenBank accession number AAC33154). CoFA and LngA share 76% amino acid sequence identity and 87% similarity. Sequences were aligned manually based on superposition of the LngA homology model with CoFA. The color scheme is as described for Fig. 1A. The secondary structure of CoFA is indicated above the sequence. LngA sequences were compared further, and variant residues are listed below the LngA sequence. (B) Superposition of the ΔN-LngA homology model (purple) on the ΔN-CoFA crystal structure (gray), showing the divergent “variable loop” that precedes strand β4. (C) Electrostatic surfaces of CoFA, LngA, and TcpA globular domains, showing the pilin face predicted to be exposed on the pilus filament. The orientation of each structure is similar to that shown in panel B. Proteins are colored according to electrostatic potential, calculated using DELPHI (39), with red representing negative charge, blue is positive, and white is uncharged (scale, −7 kT to +7 kT). (D) LngA homology model, showing that variable residues (red) are localized to the exposed LngA face.}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig6}
\caption{Heterologous expression of CoFA in \textit{Vibrio cholerae}. (A) Immunoblot showing CoFA in whole-cell culture of wild-type ETEC 31-10 and \textit{V. cholerae} (Vc) JR003, in which the gene segment encoding the mature TcpA was replaced with the corresponding cofA gene fragment within the tcp operon. CoFA from both strains ran at the expected molecular mass of the processed pilin (calculated molecular mass, 21.6 kDa) rather than the unprocessed prepilin (calculated molecular mass, 25.3 kDa). Negative controls: ETEC 31-10p, which lacks the cof operon; \textit{V. cholerae} O395, the parental strain for JR003. An immunoblot for DnaK (shown below) served as a loading control. (B) Immunoblot analysis of whole-cell supernatants, indicating the presence of intact CFA/III filaments in ETEC 31-10 but not \textit{V. cholerae} JR003.}
\end{figure}
mid housing the cof operon is present and CFA/III pili are expressed, whereas no aggregation is observed in an ETEC strain lacking the virulence plasmid or when cells are grown under non-pilus-expressing conditions. The correlation between pilus expression and ETEC aggregation suggests that it is CFA/III self-association and not the work of another protein whose gene is carried on the virulence plasmid that is responsible for aggregation, although we cannot yet rule out the latter possibility. These observations parallel those of TCP, which mediate V. cholerae aggregation, referred to as autoagglutination, as it can be visualized by eye after overnight growth in liquid culture. Our results suggest that CFA/III, like TCP, may mediate microcolony formation to facilitate colonization of the human gut.

The ETEC ∆N-CofA structure is very similar to ∆N-TcpA from V. cholerae. CofA has the characteristic type IVb pilin fold first identified in TcpA (10) but differs from TcpA in the αβ-loop, where the first half of α2 is replaced with an irregular loop containing a 310-helix. LngA is predicted to share the CofA structure, including the 310 insert, based on their sequence similarities. CofA and TcpA are more similar to each other in amino acid sequence and structure than to other type IV pili proteins of known structure: the EPEC bundlin (45) and S. Typhi PilS (58), which share the type IVb pilin fold but have markedly different αβ-loop and D-regions.

As expected from its structural similarity to TcpA, CofA fits well into the TCP pseudo-assembly resolution structure, with its 310 insert partially filling the gap between the pilin globular domains on the filament surface. Residues involved in subunit-subunit interactions are well-conserved among CofA, LngA, and TcpA, whereas amino acid differences between the ETEC type IV pili and V. cholerae TcpA are predicted to drive these filament distinctly different surfaces with respect to their chemistries and topographies. The presence of the 310 insert in CofA, together with its less-pronounced D-region, produces a smoother and less undulating filament surface for CFA/III than observed for TCP. We proposed previously that TCP self-association occurs via intercalation of protruding TcpA D-regions into cavities between subunits on adjacent TCP proteins to mediate V. cholerae autoagglutination (31). Such intercalation appears unlikely for CFA/III and Longus, but surface complementarity between adjacent filaments may be sufficient to drive ETEC aggregation. Although the CFA/III surface is predicted to have a net negative charge, we observed ETEC aggregation at and near pH 7.

Longus pili are more prevalent than CFA/III in ETEC patient isolates. A PCR-based study of 56 ETEC isolates that employed cofA- and lngA-specific primers identified 38 lngA-positive strains and only two cofA-positive strains (20). We compared published nonredundant LngA amino acid sequences and mapped the non-conserved residues to the LngA homology model (Fig. 5A and D; see also Fig. S1 in the supplemental material). Amino acid differences among LngA variants, like those of LngA, CofA, and TcpA, are localized to the exposed surface of the pilin (Fig. 5D). This may simply reflect the need to conserve residues involved in protein structure and pilus assembly, with the exposed residues being nonessential. However, since the exposed residues define the pilin filament surface, variation of these residues may profoundly alter pilus interactions with each other and with molecules in the environment, such as host receptors and molecules of the immune system. We showed previously that single amino acid substitutions on the V. cholerae TcpA surface can completely disrupt TCP-mediated autoagglutination without any apparent effects on pilin structure or pilus assembly (31). Amino acid differences among LngA variants may allow different ETEC strains to recognize their own kind through pilus-pilus interactions. It is notable that variation among LngA sequences is comparable to that between LngA and CofA (Fig. 5A). For example, ETEC E9034A LngA is 81% identical to ECOR27 LngA and is 76% identical to strain 31-10 CofA (see Fig. S1). The other proteins encoded on these pilus operons, most of which are involved in pilus assembly, are more highly conserved, both within Longus-expressing strains and between Longus and CFA/III systems (20). Thus, the cof operon can be considered a variant of the lng operon that is transmitted on its own plasmid, an idea supported by phylogenetic analyses of available cofA and lngA sequences (20).

In spite of the structural similarities between CofA and TcpA, as well as the amino acid conservation among residues involved in subunit-subunit interactions within the intact pilus, V. cholerae is not able to assemble CFA/III pili. This result suggests that one or more components of the TCP biogenesis apparatus are unable to recognize CofA. This outcome is somewhat surprising given that a number of type IVa pili have been expressed and assembled in heterologous type IVa systems (2, 5, 33, 46, 47, 56, 57), although in some cases a retraction-deficient background was required. In all of these type IVa pilus expression systems, the heterologous pilins are much less similar to the host pilins than CoFA is to TcpA, and the type IVa pilus assembly apparatus is more complex than that of type IVb (4). The inability of V. cholerae to express CFA/III pili is, however, consistent with earlier work showing that EPEC is unable to assemble V. cholerae TCP when the bfpA gene is replaced with the tcpA gene in the bfp operon (35). In this earlier study, TCP assembly likely failed in part because the TcpA signal peptide was not recognized by the bundlin prepilin peptidase. We avoided this pitfall here by inserting only the region of the cofA gene encoding the mature pilin, allowing the prepilin peptidase TcpP to recognize and remove the TcpA signal peptide. Furthermore, TcpA is more closely related to CofA than to BfpA in terms of sequence and structure. Nonetheless, we were unable to detect any pilin in our CoFA-expressing V. cholerae strain grown under either TCP- or CFA/III-inducing conditions. CoFA and TcpA are highly conserved in their N-terminal α1N segment (Fig. 1A), which serves as a membrane anchor for the pilus subunit prior to its incorporation into pilus filaments. This suggests that the TCP assembly apparatus failed to recognize the less-conserved CoFA globular domain. The inner membrane proteins required for TCP assembly, TcpE, TcpR, and TcpD (53), are not highly similar to the corresponding proteins, CofF, CofE, and CofF (37%, 0%, and 23%, respectively), which may explain their inability to recognize CoFA. Certainly the 310 insert of CoFA is a distinguishing feature of CoFA, and it may not allow interaction with TCP assembly components, but even single amino acid differences in α1 or elsewhere could potentially disrupt this specificity. The type IVb pilus assembly apparatus appears to be more specific than that of the type IVa pili, despite it being a simpler system comprising fewer components. These heterologous expression systems may provide valuable tools to dissect the pairwise interactions of the pilus assembly apparatus.

In summary, our results reveal close structural similarities between the ETEC and V. cholerae type IVb pili with respect to their subunit folds and pilus architectures and also reveal unique features that may impart distinct functions to the ETEC pili. The architectures of the pilus subunits and the pilus filament maximize
the variability among the pili to provide diverse functions within a conserved filament structure. Our results reinforce the previously observed specificity of the type IVb pilus biogenesis apparatus, which contrasts with the somewhat promiscuous type IVa pilus machinery. Our demonstration of CFA/III-mediated ETEC aggregation together with our new structure and models provides a system and a rational basis for investigating ETEC pilus interactions and type IV pilus assembly.

ACKNOWLEDGMENTS

We thank Dixon Ng for electron microscopy images, Dennis Wong, Su-maiya Islam, and Juliana Li for technical assistance, and the beamline staff at the Brookhaven National Synchrotron Light Source for their assistance in data collection.

This work was supported by operating grants from the National Institutes of Health (grant A0061051) and the Canadian Institutes of Health Research (CIHR) and by salary awards to L.C. from CIHR and the Michael Smith Foundation for Health Research.

REFERENCES