Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the CFA/III major pilin subunit CofA from human enterotoxigenic *Escherichia coli* determined at 0.90 Å resolution by sulfur-SAD phasing

Shunsuke Fukakusa,[a] Kazuki Kawahara,[b] Shota Nakamura,[c]* Takaki Iwashita,[a] Seiki Baba,[d] Mitsuhiro Nishimura,[e] Yuji Kobayashi,[e] Takeshi Honda,[c] Tetsuya Iida,[c] Tooru Taniguchi[c] and Tadayasu Ohkubo[a]

[a]Department of Pharmaceutical Sciences, Osaka University, 1-6 Yamadaoka, Suita, Osaka 565-0871, Japan, [b]Center for Research of Ancient Culture, Nara Women's University, Kita-uoya-higashi Machi, Nara City, Nara 630-8506, Japan, [c]Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan, [d]Synchrotron Radiation Research Institute, SPring-8, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan, and [e]Osaka University of Pharmaceutical Sciences, 4-20-1 Nasahara, Takatsuki, Osaka 569-1094, Japan

Correspondence e-mail: nshota@gen-info.osaka-u.ac.jp

Received 14 February 2012
Accepted 2 August 2012

PDB Reference: CofA, 3vor

CofA, a major pilin subunit of colonization factor antigen III (CFA/III), forms pili that mediate small-intestinal colonization by enterotoxigenic *Escherichia coli* (ETEC). In this study, the crystal structure of an N-terminally truncated version of CofA was determined by single-wavelength anomalous diffraction (SAD) phasing using five sulfurs in the protein. Given the counterbalance between anomalous signal strength and the undesired X-ray absorption of the solvent, diffraction data were collected at 1.5 Å resolution using synchrotron radiation. These data were sufficient to elucidate the sulfur substructure at 1.38 Å resolution. The low solvent content (29%) of the crystal necessitated that density modification be performed with an additional 0.9 Å resolution data set to reduce the phase error caused by the small sulfur anomalous signal. The CofA structure showed the $\alpha\beta$-fold typical of type IVb pilins and showed high structural homology to that of TcpA for toxin-coregulated pili of *Vibrio cholerae*, including spatial distribution of key residues critical for pilin self-assembly. A pilus-filament model of CofA was built by computational docking and molecular-dynamics simulation using the previously reported filament model of TcpA as a structural template. This model revealed that the CofA filament surface was highly negatively charged and that a 23-residue-long loop between the α1 and α2 helices filled the gap between the pilin subunits. These characteristics could provide a unique binding epitope for the CFA/III pili of ETEC compared with other type IVb pili.

1. Introduction

Enterotoxigenic *Escherichia coli* (ETEC) can cause diarrhoea in children and travellers in developing countries (Qadri *et al.*, 2005). Adherence of ETEC to host cell surfaces, colonization of the small intestine and elaboration of heat-labile and/or heat-stable enterotoxins are essential steps in ETEC pathogenesis (Fleckenstein *et al.*, 2010). In human ETEC, the adherence and colonizing abilities depend on the presence of colonization factor antigens (CFAs), which form pili (or fimbriae) that attach to the target cell surface.

Several types of CFAs and putative colonization factors (PCFs) have been identified. These antigens are categorized as CFA/I, CFA/II, CFA/III, CFA/IV, PCFO148, PCFO159, PCFO166 *etc.* on the basis of their antigenic specificity or the N-terminal amino-acid sequence of the major pilin subunit (Gaastra & Svennerholm, 1996). Previously, we determined the nucleotide sequence of the whole region required for the formation of CFA/III of human ETEC. We identified a cluster

of 14 genes, including *cofA* and *cofP* (Taniguchi *et al.*, 1995, 2001). The formation of CFA/III pili required processing of the N-terminal signal sequence of CofA, a major pilin subunit of CFA/III, by a prepilin peptidase (CofP), which is similar to the assembly process of other type IV pili (Craig *et al.*, 2004).

Type IV pilins form long thin fibrils (each 60–90 Å in diameter) that mediate diverse biological functions, including adhesion, surface motility, microcolony formation, biofilm formation and DNA uptake (Dhakal *et al.*, 2009; Higashi *et al.*, 2011). Each pilin subunit is characterized by an unusual N-methylated N-terminus, a conserved hydrophobic N-terminal fibre-forming domain and a carboxy-terminal disulfide bond. Type IV pilins are classified into two subgroups, types IVa and IVb, primarily based on their signal sequence length and the size of the mature protein (Pelicic, 2008). In particular, the N-methylated N-terminal residue of mature type IVa pilin is phenylalanine, whereas that of mature type IVb pilin is methionine, leucine or valine (Fig. 1; Craig *et al.*, 2004). The N-terminal 30-residue segment of mature CofA produced after processing of the N-terminal signal sequence is highly hydrophobic. This segment shows high sequence homology, including the conserved N-methylated N-terminal residue, to other type IVb pilins, such as TcpA for toxin-coregulated pili (TCP) of *Vibrio cholerae*, BfpA for bundle-forming pili (BFP) of enteropathogenic *E. coli* (EPEC), PilS for *Salmonella* pili of *Salmonella* Typhi and LngA for the long pilus (longus) of ETEC (Fig. 1; Girón *et al.*, 1997; Gómez-Duarte *et al.*, 1999; Taniguchi *et al.*, 1994, 1995). Consequently, CofA for CFA/III of ETEC can be categorized as a type IVb pilin.

Although an understanding of the assembly mechanisms of type IVb pilin subunits and the structures of type IVb pili is crucial for the development of novel vaccines against bacterial pathogens, these concepts are incompletely understood. Recently, utilizing X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, the structures of three type IVb pilin subunits, TcpA of *V. cholerae*, BfpA of EPEC and PilS of *Salmonella* Typhi, were independently determined using N-terminally truncated (by ∼28 amino acids) constructs (Craig *et al.*, 2003; Xu *et al.*, 2004; Ramboarina *et al.*, 2005). These structures revealed that the type IVb pilin structures share a common overall architecture including an N-terminal α-helix and a central antiparallel β-sheet, both of which form a hydrophobic structural core.

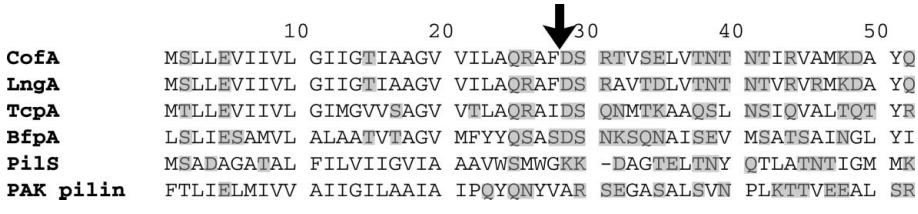


Figure 1
Amino-acid sequence alignment of the N-terminal α-helices of type IV pilins. The N-terminal sequence (∼50 amino acids) of CofA of ETEC is aligned with those of other type IVb pilins, including LngA of ETEC, TcpA of *V. cholerae*, BfpA of EPEC and PilS of *Salmonella* Typhi. The N-terminal sequence of PAK pilin (categorized as type IVa) is shown for comparison. Polar residues are shaded in grey. The black arrow indicates the truncated site used for construction of the N-terminally truncated version of CofA.

Beyond the conserved structural core, the pilin structures vary substantially in two characteristic regions: the $\alpha\beta$-loop, which is situated between the C-terminal α-helix and the central β-sheet, and the D-region, which is located between two conserved disulfide-bonded cysteines (Li *et al.*, 2008). Because these regions are solvent-exposed and cover much of the protein surface, they could be responsible for pilus functions such as pilus–receptor and pilus–pilus interactions. For example, it has been reported that PilS of *Salmonella* Typhi binds the cystic fibrosis transmembrane conductance regulator (CFTR), an epithelial cell receptor for *Salmonella* pili of *Salmonella* Typhi, and that both the $\alpha\beta$-loop and the D-region are responsible for its binding (Xu *et al.*, 2004; Balakrishna *et al.*, 2009). For TcpA, residues that are important for pilin self-assembly reactions and subsequent pilus–pilus interactions have been identified in these regions (Kirn *et al.*, 2000). These examples emphasize the functional importance of the solvent-exposed regions in bacterial pathogenesis. The diversity among type IVb pilin subunits from different bacteria, especially in the sequence and structure of the solvent-exposed regions, implies that the pili vary functionally to attain efficient colonization under specific environments and, possibly, to escape antibody-mediated immunity.

Although studies have revealed the structural core of type IV pilin, it has been extremely difficult to determine the atomic resolution structure of the full-length type IVb pilin subunit. This difficulty arises from the ability of the conserved hydrophobic N-terminal segment to form a protruding α-helix and to enhance self-assembly to form protein aggregates. In this context, promising approaches have been applied to obtain such information, including cryo-electron microscopy (cryo-EM) and X-ray fibre-diffraction studies of wild-type type IV pili separated from the bacteria coupled with mutational analyses and computational docking (Kirn *et al.*, 2000; Parge *et al.*, 1995; Campos *et al.*, 2011). Hydrogen/deuterium-exchange mass spectrometry (DXMS), which detects changes in protein dynamics induced by protein–protein interactions, has been shown to be effective for identifying the interaction interface and has recently been used to refine the pilus-filament model (Li *et al.*, 2008). These techniques have been used to elaborate several plausible pilus-filament models of TcpA of *V. cholerae* and BFP of EPEC. The resulting models provide some insight into the molecular basis of pilus assembly and display the binding epitopes of the pili, although such data for other type IVb pili remain very limited (Li *et al.*, 2008; Ramboarina *et al.*, 2005).

In the present study, in order to reveal the structure and assembly mechanism of CFA/III of ETEC in detail, we determined the crystal structure of an N-terminally truncated version of its major pilin subunit, CofA, at an extremely high resolution (0.90 Å) by single-wavelength anomalous diffraction (SAD) phasing utilizing five intrinsic S atoms in the

Table 1
Summary of data-collection, phasing and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Native	SAD
Data-collection statistics		
Beamline	BL38B1, SPring-8	BL38B1, SPring-8
Detector	ADSC Quantum 315	ADSC Quantum 315
Wavelength (Å)	0.9000	1.5000
Exposure time (s)	3.3	3.0
Crystal-to-detector distance (mm)	75.0	75.0
Oscillation angle (°)	1.0	1.0
Total oscillation range (°)	180	720
Resolution (Å)	30.86–0.90 (0.95–0.90)	47.64–1.38 (1.42–1.38)
Space group	$P2_1$	$P2_1$
Unit-cell parameters (Å, °)	a = 33.82, b = 47.71, c = 42.45, β = 107.52	a = 33.82, b = 47.64, c = 42.53, β = 107.57
No. of unique reflections	91809	51931
$R_{\rm merge}$†	0.050 (0.170)	0.041 (0.226)
Completeness (%)	96.4 (94.9)	99.5 (99.6)
Wilson B factor (Å^2)	3.2	8.2
$\langle I/\sigma(I)\rangle$	18.6 (7.1)	32.1 (6.3)
Multiplicity	3.6 (3.5)	6.8 (4.7)
Anomalous signal to noise‡		1.06 (0.79)
Phasing statistics		
No. of sites		5 sulfurs
Figure of merit (FOM)		0.833
Refinement data statistics		
Resolution range (Å)	30.86–0.90 (0.91–0.90)	
No. of reflections used in refinement	91733 (2850)	
Overall average B factor (Å^2)	4.5	
No. of protein atoms	3189	
No. of waters	237	
Final $R_{\rm work}$	0.120 (0.146)	
Final $R_{\rm free}$§	0.130 (0.158)	
Rotamer outliers (%)	1.1	
Clash score	6.42	
R.m.s. deviations from ideal values¶		
Bonds (Å)	0.01	
Angles (°)	1.54	
Ramachandran plot analysis		
Most favoured regions (%)	97.7	
Disallowed regions (%)	0.0	

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity of the ith measurement of reflection hkl and $\langle I(hkl)\rangle$ is the mean intensity of reflection hkl calculated after scaling. ‡ The anomalous signal to noise, $\langle F^{\pm}\rangle/\langle\sigma(F)\rangle$, is the ratio of the anomalous differences to the noise. § $R_{\rm free}$ was calculated for a random set of 5% of reflections that were not used in the refinement. ¶ Ideal values are derived from the geometry library of Engh & Huber (2001).

native protein. The resulting structure showed the $\alpha\beta$-fold typically found in the type IVb pilin subunit and showed high structural homology with that of TcpA of *V. cholerae*, including several residues that are critical for pilus-filament formation. Based on the sequence and structural similarities, we propose a pilus-filament model for CofA. The model reveals that the pilus filament has unique surface properties compared with other type IVb pilus family members.

2. Materials and methods

2.1. Expression and purification of the N-terminally truncated version of CofA

Based on the results of a sequence comparison with other type IVb pilins (Fig. 1), a construct of CofA was designed in which a conserved N-terminal hydrophobic segment (1–28) was truncated to solubilize the protein. The DNA sequence encoding CofA was PCR-amplified from plasmid pTT240, which contains the CFA/III gene cluster (GenBank ID AB049751). The forward and reverse primers were 5′-ACGA**GGATCC**GACTCACGTACTGTTTCTGAATTG-3′ (the *Bam*HI site is shown in bold) and 5′-ACCTGTAT**AAGCTT**ATCTTATTAACGGCTCGCCAAAG-3′ (the *Hin*dIII site is shown in bold), respectively. The amplified PCR product was digested with the restriction enzymes *Bam*HI and *Hin*dIII and ligated into a modified pET32b vector (Merck Biosciences, Germany).

E. coli expression strain Origami B (DE3) (Novagen, USA) cells were transformed with the recombinant plasmid for overexpression of the N-terminally truncated CofA, which was fused with thioredoxin and had a His_6 tag at its N-terminal end. The cells were grown in LB medium containing ampicillin (50 µg ml^{-1}) until the optical density of the culture at 600 nm and 310 K reached approximately 0.6. The culture was induced with 0.1 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated for 6 h at 290 K. The cells were centrifuged at 6000*g* for 15 min. The cell pellet was resuspended in lysate buffer consisting of 10 m*M* phosphate-buffered saline (PBS) pH 7.4, 0.1% Triton X-100, 0.2 mg ml^{-1} lysozyme. The suspension was sonicated on ice for 1 min with a Branson Sonifier 150 (Branson Ultrasonics, USA) with a 1 min interval between each sonication. This procedure was repeated ten times. The lysate was centrifuged at 35 000*g* for 40 min at 277 K.

The supernatant was loaded onto a nickel-chelating column (HiTrap Chelating HP, GE Healthcare, USA) that had been pre-equilibrated with lysis buffer (10 m*M* PBS pH 7.4, 10 m*M* imidazole). CofA containing a His_6 tag was eluted with a linear gradient from 0 to 500 m*M* imidazole in 10 m*M* PBS pH 7.4 buffer. The protein was eluted at imidazole concentrations of between 40 and 100 m*M*. Fractions containing CofA, as determined by SDS–PAGE, were pooled and dialyzed against buffer (10 m*M* PBS pH 7.4) for 12 h. Finally, the fusion tag was cleaved with thrombin by overnight incubation at 293 K. The dialyzed sample was again purified using a nickel-chelating column. The cleaved CofA protein contained two extra residues (Gly-Ser) at the N-terminus.

The protein solution was concentrated using Amicon YM-10 membranes (Amicon, USA). The concentrated protein sample was further purified using a size-exclusion column (Superdex 75 26/60, GE Healthcare, USA) that had

been pre-equilibrated with buffer (20 m*M* Tris–HCl pH 7.5). Finally, the purified CofA was checked by SDS–PAGE and concentrated to 30 mg ml^{-1} in 20 m*M* Tris–HCl pH 7.5.

2.2. Crystallization of CofA

The crystallization conditions for CofA were initially screened by the sitting-drop vapour-diffusion method at 293 K with Crystal Screen and Crystal Screen 2 (Hampton Research, USA), JB Screen kits (Jena Bioscience, Germany) and Wizard Screen I and II (Emerald BioStructures, USA). Each crystallization drop was prepared by mixing 2 µl protein solution and 2 µl reservoir solution and was equilibrated against 400 µl reservoir solution. After a few days, small crystals of CofA were obtained under several conditions in which polyethylene glycol (PEG) 1500 or PEG 4000 was used as a precipitant agent. These conditions were further optimized by varying the pH and the concentration of the protein and precipitants. Optimal crystals for X-ray diffraction studies were grown in a week from drops consisting of 7.5 mg ml^{-1} CofA with 26% PEG 4000 and 100 m*M* acetate buffer pH 3.8. The crystals used for data collection had a rod-like shape, with typical dimensions of 0.1 × 0.1 × 0.5 mm.

2.3. Diffraction data collection and processing

Single crystals of CofA mounted in a Litho loop (Protein Wave Co., Japan) were soaked in cryoprotectant solution consisting of 17% glycerol, 26% PEG 4000, 100 m*M* acetate buffer pH 3.8 for 10 min prior to X-ray diffraction analysis. The cryoprotectant solution in the loop was removed as far as possible and the mounted crystal was flash-cooled in a cold nitrogen-gas stream. X-ray diffraction data sets were collected at 100 K under a cold nitrogen stream and were recorded using an ADSC Quantum 315 or Quantum 315r CCD detector on beamline BL38B1 at SPring-8, Hyogo, Japan.

Two types of diffraction experiment were performed for data collection. A native data set was collected at a wavelength of 0.9 Å with a total oscillation range of 180°, an oscillation angle of 1.0° and a crystal-to-detector distance of 75 mm. The SAD data sets were collected at wavelengths ranging from 1.3 to 1.7 Å with a total oscillation range of 720° each, an oscillation angle of 1.0° and a crystal-to-detector distance of 75 mm (Table 1 and Supplementary Table S1[1]). The native and SAD diffraction data sets were collected independently from different crystals in order to reduce any possible radiation damage to the crystal.

Single crystals of CofA diffracted very well to a resolution of 0.90 Å at a wavelength of 0.9 Å and belonged to space group $P2_1$, with unit-cell parameters a = 33.82, b = 47.71, c = 42.45 Å, β = 107.52°. The Matthews coefficient was estimated to be 1.73 Å^3 Da^{-1}, with an estimated solvent content of 29%, assuming the presence of one CofA molecule in the asymmetric unit (Matthews, 1968). The Matthews coefficient and solvent content were calculated with programs from the *CCP*4 suite (Winn *et al.*, 2011).

To maximize the anomalous signal of S atoms and to reduce undesired absorption effects of the solvent, a wavelength of 1.5 Å was used for sulfur-SAD phasing. This process resulted in a medium-resolution (1.38 Å) data set with similar unit-cell parameters to those of the native data set.

All of the diffraction data were integrated and scaled with the *XDS* program package (Kabsch, 2010). The anomalous signal was analyzed with the program *XSCALE* (Kabsch, 2010). The data-collection and processing statistics are summarized in Table 1. The number of reflections, completeness, $R_{\rm merge}$ and $I/\sigma(I)$ values of the high-resolution (0.90 Å) native data set in each resolution shell are also given in Supplementary Table S2.

2.4. Structure determination of CofA

The *SHELXC*/*D*/*E* programs (Sheldrick, 2010; Pape & Schneider, 2004) were used for sulfur-substructure determination, initial phasing and phase improvement by density modification. By combining the use of medium-resolution (1.38 Å) and high-resolution (0.9 Å) data collected at wavelengths of 1.5 and 0.9 Å, respectively, an electron-density map of excellent quality was obtained which was capable of determining each atomic position unambiguously (see §3 for details). The overall mean figure of merit (FOM) of the obtained phase was 0.833.

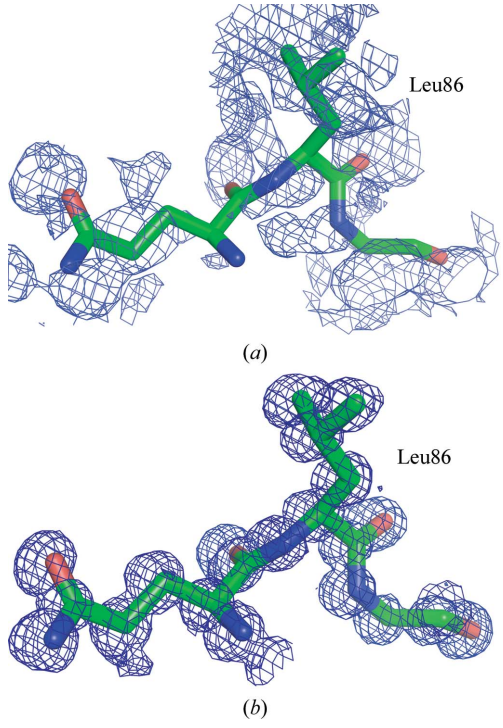


Figure 2
$2mF_{\rm o} - DF_{\rm c}$ electron-density map covering residues 85–87 contoured at 0.5 e Å^{-3}. The map in (*a*) was calculated using the phase obtained from diffraction data collected at a wavelength of 1.5 Å and that in (*b*) was calculated using the improved phase in a density-modification procedure with high-resolution diffraction data collected at a wavelength of 0.9 Å.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: KW5045). Services for accessing this material are described at the back of the journal.

Model building was initially performed using diffraction data limited to 2.0 Å resolution. Of the 182 residues of CofA, 166 residues were automatically built at this stage using the program *phenix.autobuild* from the *PHENIX* program suite (Adams *et al.*, 2002). The model was subjected to iterative rounds of refinement while gradually increasing the resolution from 2.0 to 0.9 Å using the program *phenix.refine*. This process was followed by manual correction with *Coot* (Emsley *et al.*, 2010). In the final round of refinement, H atoms and alternative conformations were included and anisotropic temperature-factor parameters for non-H protein atoms were refined. The geometry of the final model of CofA was checked with *MolProbity* (Chen *et al.*, 2010). The phase-determination and refinement statistics are listed in Table 1. Figures were drawn with the *PyMOL* molecular-graphics system (v.1.3; Schrödinger LLC). The final coordinates of the CofA structure were deposited in the Protein Data Bank (PDB entry 3vor; Berman *et al.*, 2000).

2.5. Modelling the pilus of CofA

The filament model of CofA containing 18 subunits was built on the basis of homology modelling and computational docking with *MODELLER* 9.10 (Eswar *et al.*, 2006) and *Coot* (Emsley *et al.*, 2010). The crystal structure of CofA was superposed on each of the pilin subunits in the TcpA filament model (PDB entry 1orq) previously reported by Craig *et al.* (2003). For each case, the backbone C^{α} atoms of the two structures superposed well, with a root-mean-square (r.m.s.) deviation value of around 2.4 Å. However, many residues (681 residues in 18 subunits) at the protein surface interacted sterically (5836 contacts) with those of neighbouring subunits because the interatomic distance was shorter than the van der Waals (vdW) interaction (the sum of the vdW radii of each atom − 0.4; Jackson & Sternberg, 1995), except those involving hydrogen-bond interaction(s). Thus, the filament model was subsequently refined using a molecular-dynamics module with a simulated-annealing protocol and CHARMM22 energy terms in combination with the conjugate-gradient module of *MODELLER* 9.10 (Eswar *et al.*, 2006). Because the N-terminal segment (1–28) of CofA was not included in the construct and no full-length crystal structure has been reported for the type IVb pilin subunit, the N-terminal hydrophobic segment of CofA was modelled using the full-length PAK pilin structure of type IVa pilin (PDB entry 1oqw; Craig *et al.*, 2003) as a structural template. Sequence alignment of CofA and PAK pilin using *ClustalW* (Larkin *et al.*, 2007) showed high homology at the N-terminal 50 residues, with a sequence identity of 26% and a similarity of 82%, even though the N-methylated N-terminal residues of the mature pilin subunits were different (Met in CofA and Phe in PAK pilin; Fig. 1). As expected, part of the N-terminal α-helix (37–50) of the CofA structure superposed well with that of the PAK pilin structure. Thus, the only other remaining residues (1–36) of CofA were modelled using a method implemented in *MODELLER* 9.10 (Eswar *et al.*, 2006). To stabilize the interactions among the N-terminal α-helices during the course of the simulated-annealing process, the interaction between the positively charged N-terminal amine of Met1 and the negatively charged Glu5 side chain was constrained. The interaction was made manually by adjusting the orientation of the N-terminal α-helices.

After the simulated-annealing procedure, an additional energy-minimization process with no structural constraints was undertaken until the energy of the system converged to 100 kJ mol^{-1} nm^{-1}. The energy-minimization process was performed with the quasi-Newton method in the *GROMACS* 4.5.4 program package (Hess *et al.*, 2008) and the energy calculation was performed with the AMBER99 energy terms under the NPT ensemble (*i.e.* constant pressure and temperature) and periodic boundary conditions. Simple point charge (SPC) water molecules were chosen for the water model and sodium counter-ions (54 Na^{+} ions) were added to obtain an electrically neutral system. At each refinement step, atomic contacts between the subunits were assessed with the program *CONTACT* (Winn *et al.*, 2011).

3. Results and discussion

3.1. Structural determination of CofA by sulfur-SAD

Removal of the N-terminal hydrophobic segment (1–28) of CofA yielded a highly soluble protein that permitted the successful crystallization of the pilin subunit, similar to other type IVb pilins (Hazes *et al.*, 2000). The crystal diffracted very well to a resolution of 0.90 Å using a wavelength of 0.90 Å. The solvent content of the crystal was relatively low (calculated as 29%), with one protein molecule in the asymmetric unit (Matthews, 1968).

An initial attempt at phase determination by the molecular-replacement method using previously reported structures of type IVb pilins did not yield a clear solution. Using the

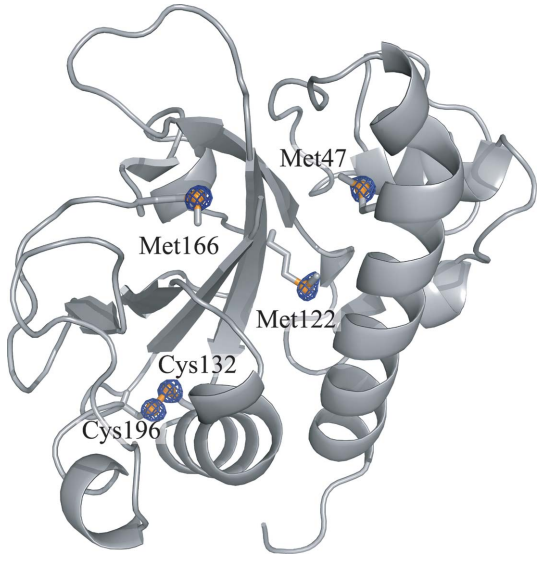


Figure 3
Crystal structure of CofA superimposed with the five sulfur sites located by *SHELXD*. An anomalous difference Fourier map contoured at 2.0σ and calculated at 0.90 Å resolution with the program *FFT* is also shown.

program *Phaser* with a polyalanine model of TcpA (PDB entry 1orq) in which all of the side chains were pruned, we produced one feasible solution with a translation-function *Z*-score (TFZ) of 3.8, a rotation-function *Z*-score (RFZ) of 6.0 and a log-likelihood gain (LLG) of 26.0 (McCoy *et al.*, 2007). However, the resulting electron density was poor and it was difficult to build a model without ambiguity. Thus, the structure was determined by SAD phasing utilizing the anomalous signal of the S atoms from the native protein.

The N-terminally truncated version of CofA contained 182 residues, including three methionine and two cysteine residues. The expected anomalous signal ($\langle\Delta F^{\pm}\rangle/\langle F\rangle$) from the five S atoms at a wavelength of 1.5 Å was 0.54% (Hendrickson & Teeter, 1981), which is comparable to the current limit (0.48%) of phasing by sulfur-SAD (Wang *et al.*, 2006). Therefore, the data-collection parameters, especially the wavelength, were carefully adjusted to reduce the undesired absorption effects of the solvent and to enhance the intrinsically small intensity of the anomalous sulfur signal. Diffraction data were collected at wavelengths of 0.9–1.7 Å on beamline BL38B1 at SPring-8, Hyogo, Japan. A meaningful anomalous signal was only observed at a wavelength of 1.5 Å (Table 1, Supplementary Table S1 and Fig. S1). Using these data, we successfully located the sites of all five S atoms using the program *SHELXD* (Schneider & Sheldrick, 2002). However, subsequent phase calculation followed by density modification with the programs *SHELXE* (Sheldrick, 2002; Cowtan & Zhang, 1999), *MLPHARE* (Winn *et al.*, 2011), *OASIS*-2006 (Wang *et al.*, 2004) and *SHARP* (Bricogne *et al.*, 2003) failed to produce an interpretable electron-density map (Fig. 2*a*). This result was possibly owing to the low solvent content of the crystal (29%), which arose from its tight molecular packing.

The solvent content is reported to be an important factor in successful sulfur-SAD phasing with a weak anomalous signal (Ramagopal *et al.*, 2003). Although necessary to reduce the phase error, it is difficult to apply a density-modification procedure with relatively low-resolution data (in this case, 1.38 Å) when

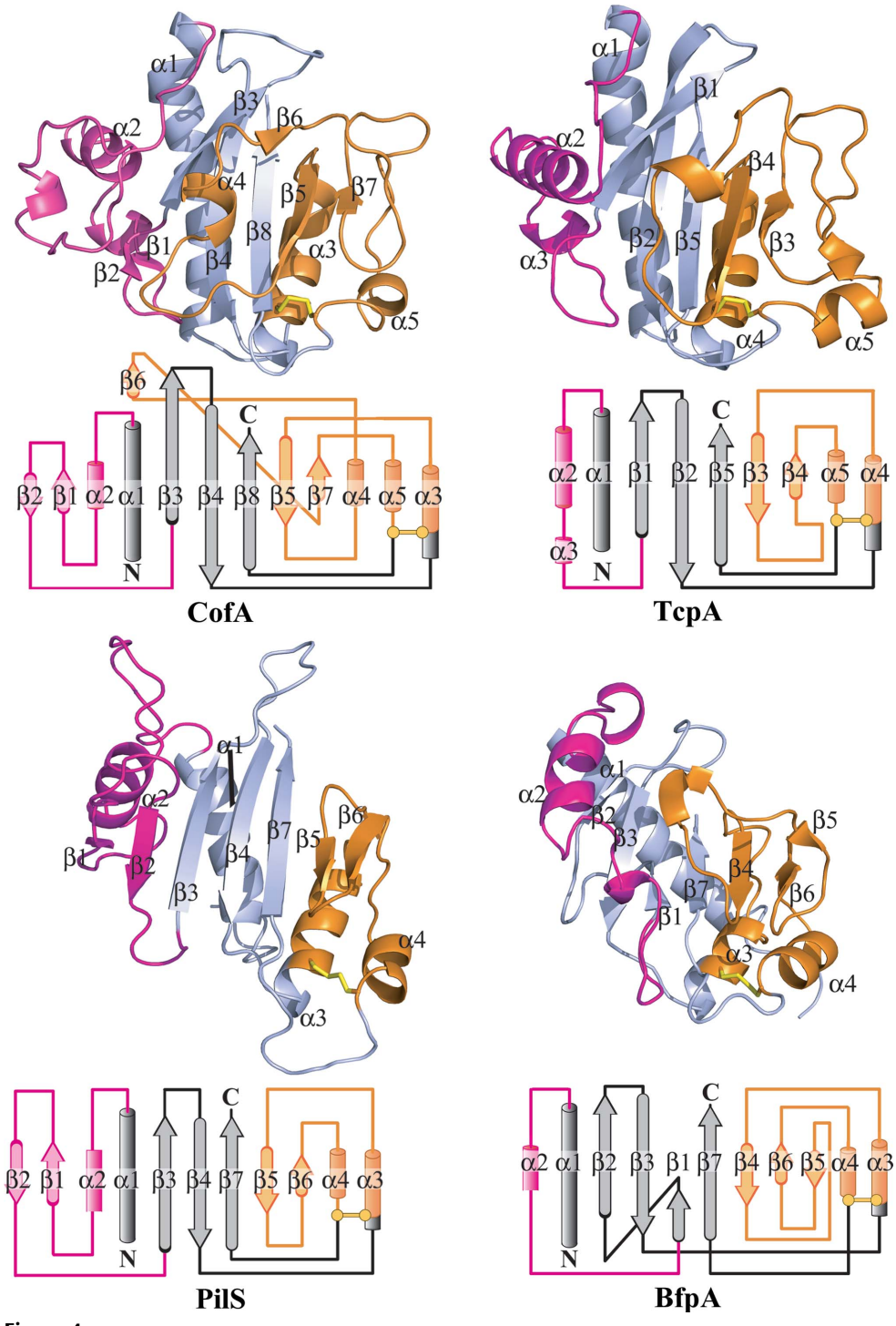


Figure 4
Structural comparison between CofA and other type IV pilins. Crystal structures of previously determined type IVb pilins, TcpA (PDB entry 1oqv), PilS (PDB entry 1q5f) and BfpA (PDB entry 1zwt), are shown. In each case, the tertiary-structural model (above) and topology diagram (below) are provided. The D-region and $\alpha\beta$-loop are represented in orange and magenta, respectively. The hydrophobic structural core is shown in grey. Disulfide bonds are represented by a yellow stick model.

Table 2
Sequence and structural similarity between CofA and other type IVb pilins.

	TcpA	PilS	BfpA
Sequence†			
Identity (%)	28.0	17.0	14.0
Similarity (%)	68.0	49.0	51.0
Structure‡			
R.m.s. deviation (core) (Å)	1.90	2.84	2.74

† The values for amino-acid sequence identity and similarity between two structures were calculated based on the sequence-alignment result obtained using the program *ClustalW*2 (Larkin *et al.*, 2007). ‡ The r.m.s. deviation value was calculated by superimposing each structure on the CofA structure with the *SSM* superposition routine (Krissinel & Henrick, 2004) implemented in the program *Coot* (Emsley *et al.*, 2010).

the solvent content is very low (<30%). Therefore, density modification was performed again using the program *SHELXE* with the sphere-of-influence algorithm (Sheldrick, 2008) and high-resolution (0.90 Å) native data. This process generated a well defined and easily interpretable electron-density map (Fig. 2*b*). Approximately 90% of the residues in the monomer were automatically built with the program *phenix.autobuild* in the *PHENIX* program package (Adams *et al.*, 2002). The remaining residues (including several alternative conformations) were built manually. The final coordinates included all of the residues in our designed CofA construct, with final R_{work} and R_{free} values of 0.120 and 0.130, respectively (Table 1). The r.m.s. deviations of the bond lengths and bond angles of the CofA structure from ideal values were 0.01 Å and 1.54°, respectively (Engh & Huber, 2001). An anomalous difference Fourier map calculated with phases from the final CofA model clearly showed five peaks (Fig. 3).

3.2. Overall structure of CofA

The overall structure of CofA determined at 0.90 Å resolution was that of the typical $\alpha\beta$-fold commonly found in other reported type IV pilin structures (Parge *et al.*, 1995) constituting a central six-stranded β-sheet surrounded by five α-helices, including two 3_{10}-helices (Figs. 3 and 4). The structural core of CofA, which was flanked by two characteristic regions (the $\alpha\beta$-loop and D-region), was formed by an antiparallel β-sheet composed of three long β-strands (β3, β4 and β8) interacting with an N-terminal α-helix mainly by hydrophobic interactions.

The $\alpha\beta$-loop connecting the α1 helix and the central β-sheet was composed of 49 amino acids (55–103) and included one α-helix (α2) in the middle of the loop flanked by two 3_{10}-helices. A short β-sheet consisting of two β-strands (β1 and β2) was observed at the loop near the C-terminal side of the α2 helix. A D-region of about 65 amino acids (132–196) on the other side of the molecule connected three α-helices, α3, α4 and α5, which were linked by a conserved disulfide bridge between Cys132 and Cys196 that stabilized the structure (Li *et al.*, 2008). The D-region interacted with the central β-sheet, forming a six-stranded mixed parallel ($\beta6\rightarrow\beta3$) and antiparallel ($\beta4\rightarrow\beta8\rightarrow\beta5\rightarrow\beta7$) β-sheet with a barrel-like shape. The $\alpha\beta$-loop and D-region were highly solvent-exposed, covering almost the entire structural core of CofA. This characteristic suggested functional significance of these regions, such as in pilus assembly and pilus–pilus interactions.

3.3. Comparison of the structure of CofA with those of other type IVb pilins

The $\alpha\beta$-loop and D-region of type IVb pilins show substantial differences in sequence, size and conformation depending on the bacterial origin. These differences provide the functional variety of the pilin subunit (Craig *et al.*, 2003; Xu *et al.*, 2004; Ramboarina *et al.*, 2005). Three structures of type IVb pilin subunits, TcpA of *V. cholerae*, BfpA of EPEC and PilS of *S.* Typhi, have been reported (Fig. 4 and Table 2). A structural comparison clearly revealed that the structural core, which is composed of an antiparallel β-sheet and an N-terminal α-helix, is well conserved among the pilins, whereas notable differences were found in the $\alpha\beta$-loop and the D-region (Craig *et al.*, 2003; Xu *et al.*, 2004; Ramboarina *et al.*, 2005). In the D-region, the length of the loops and their topology are basically the same among the pilins, except in the case of PilS from *Salmonella* Typhi. PilS has the shortest D-region among the type IVb pilin structures and the central five-stranded antiparallel β-sheet of PilS from *Salmonella* Typhi is highly exposed to solvent. The corresponding regions in other pilins are replaced by long loops, including one or two short secondary elements covering a large portion of the protein surface.

The structure of the $\alpha\beta$-loop is more divergent depending on the pilin structure. In particular, the $\alpha\beta$-loops of CofA and

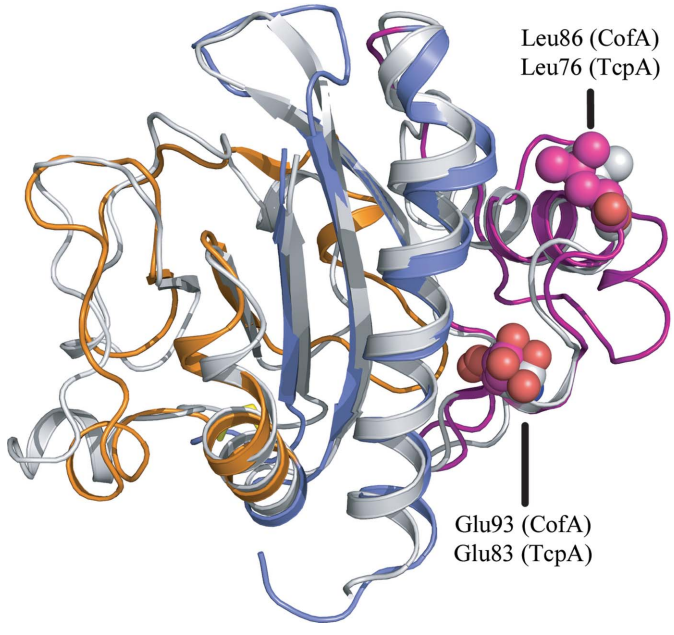
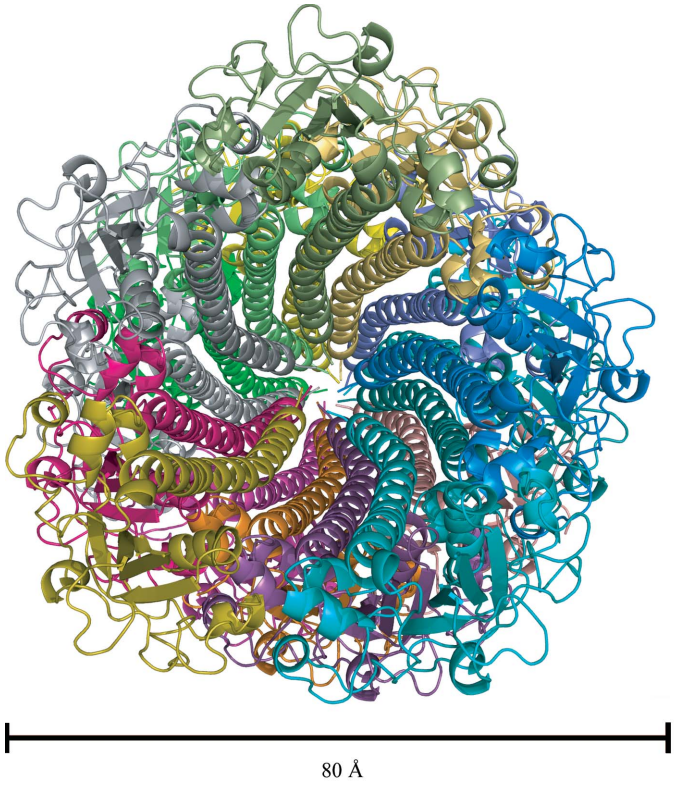


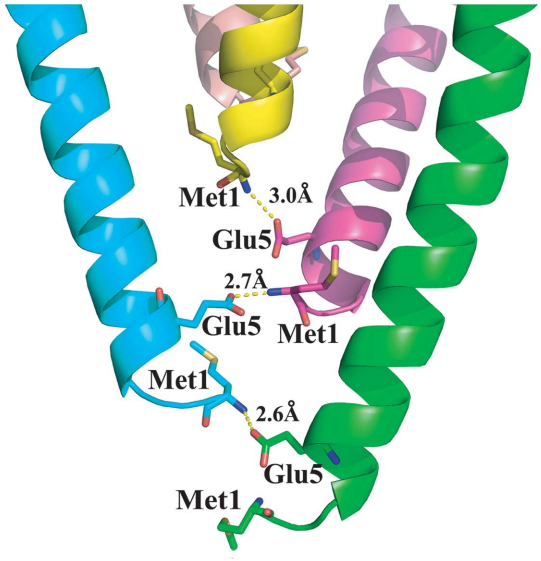
Figure 5
Superposition of the CofA and TcpA structures. The structures of CofA and TcpA are shown as cartoon models. The hydrophobic structural core, $\alpha\beta$-loop and D-region of the CofA structure are coloured blue, magenta and orange, respectively. The TcpA structure is coloured grey. Two key residues (Leu76 and Glu83 in TcpA and Leu86 and Glu93 in CofA) that are considered to be critical for pilin self-assembly are shown as spheres.

PilS uniquely contain an extra pair of parallel β-sheets which are composed of two short β-strands (β1 and β2) that rigidify and stabilize the C-terminal region of the $\alpha\beta$-loop structure. Although the binding partner of CofA is not well characterized, previous studies have shown that the PilS pilin binds to a peptide segment, GRIIASYDPDNKEER. This segment corresponds to the first extracellular domain of human CFTR at a region in close proximity to the extra pair of antiparallel β-sheets, revealing the importance of this solvent-exposed region for substrate recognition (Balakrishna *et al.*, 2009). In

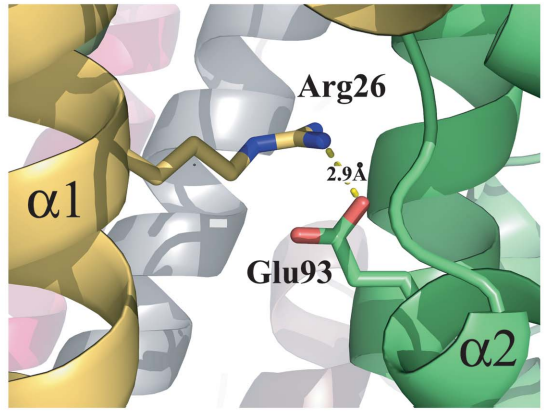


(*a*)

(*b*)



(*c*)



(*d*)

Figure 6
Filament model of CofA. (*a*) Top view of the CofA filament model. The filament diameter is approximately 80 Å. (*b*) Side view of the CofA filament model. The CofA model contains 18 molecules. (*c*) Close-up of the interaction between Met1 and Glu5 in the N-terminal α-helix. (*d*) Close-up of the intersubunit interaction between Arg26 and Glu93. (*e*) Close-up of the α1–α2 and α5–α2 contacts. Leu86 at the end of the α2 helix interacts with Ala189 and Ile192 on the α5 helix.

contrast, the corresponding counterpart of TcpA forms an extended loop. BfpA also forms a β-strand (β1) in the corresponding region; however, the strand is incorporated into the central β-sheet, forming a seven-stranded parallel and antiparallel β-sheet.

In spite of their structural differences, in all of the type IVb pilin structures that have been reported to date the $\alpha\beta$-loop contains a short α-helix (α2) on one edge of the central β-sheet. The α2 helix is connected to the N-terminal α1 helix by a relatively short loop (typically ~15 residues). In contrast, in CofA a 23-residue loop (55–77) was inserted between the α2 helix and the N-terminal α1 helix (Figs. 4 and 6).

3.4. Pilus model of CofA

Interestingly, as judged from superposition of the structures using the backbone C^{α} atoms, the structural core of CofA was almost identical to that of TcpA from *V. cholerae*, with an r.m.s. deviation of 2.16 Å (Figs. 4 and 5). TcpA is one of the best-studied examples of type IVb pilus formation. A plausible filament model has recently been proposed and elaborated *via* multiple techniques (*e.g.* mutational studies, cryo-EM reconstruction, X-ray crystallography, DXMS and computational docking; Li *et al.*, 2008). In the refined model of TcpA, five specific residues (Met1, Glu5, Arg26, Leu76 and Glu83) stabilize the subunit–subunit interactions in two regions: one at the N-terminal α-helices and the second between the $\alpha\beta$-loop and the D-region (Li *et al.*, 2008). Remarkably, a sequence alignment showed that all of the key residues were also conserved in CofA and the two structures could be superposed (Figs. 1 and 5). Moreover, residues Leu86 and Glu93 of CofA occupied almost the same spatial positions as the corresponding residues, Leu76 and Glu83, of TcpA in the $\alpha\beta$-loop. We speculated that the mechanisms of pilus assembly for these two pilins were similar, even though their overall amino-acid sequence identity was marginal (~30%).

Based on the above-mentioned similarities, we built a pilus model of CofA by homology modelling. After constructing the N-terminal hydrophobic segment (1–28), we performed molecular-dynamics simulation using the publicly available TcpA filament model obtained based on the crystal structure of TcpA (PDB entry 1oqv). Our initial simple superimposition of the CofA structure on each pilin subunit of the TcpA filament model resulted in severe steric interference at the subunit–subunit interface owing to structural differences at the solvent-exposed region (Supplementary Fig. S2). To alleviate this interference, structural refinements were performed. A total of 44 cycles of molecular-dynamics simulation were performed with the simulated-annealing protocol (each consisting of 200 ps of molecular dynamics with a time step of 10 fs), followed by energy minimization utilizing a conjugate-gradient method. The hydrophobic N-terminal segment appears to be critical for subunit assembly and the electrostatic interaction between the N-methylated N-terminal amine of Met1 and polar atoms of Glu5 seems to play an important role in subunit bundling (Li *et al.*, 2008). Accordingly, we constrained this interaction during the simulated-annealing process. The structure-refinement procedure successfully eliminated all of the observed steric interference (Supplementary Fig. S2). No meaningful structural changes were observed when additional energy minimizations were performed with the *GROMACS* software package without structural constraints at the N-terminus.

As expected from their high similarity in structure and size (208 residues for CofA and 199 residues for TcpA), the filament model of CofA was visually similar to that of TcpA (Figs. 6*a* and 6*b*). The filament model of CofA showed a left-handed three-start helical arrangement of each pilin subunit, with a diameter of about 80 Å. This diameter is comparable to that of pili found in ETEC (~70 Å) and slightly larger than those of typical type IVa pili (~60 Å) (Gaastra & Svennerholm, 1996; Touhami *et al.*, 2006). In our CofA filament model, each subunit interacted at the N-terminal α-helices and at a region between the $\alpha\beta$-loop and the D-region (Fig. 6). The critical interaction network stabilizing the N-terminal α-helices is shown in Fig. 6(*c*). In addition to the electrostatic interaction between Met1 and Glu5, a salt bridge was formed between Arg26 in the N-

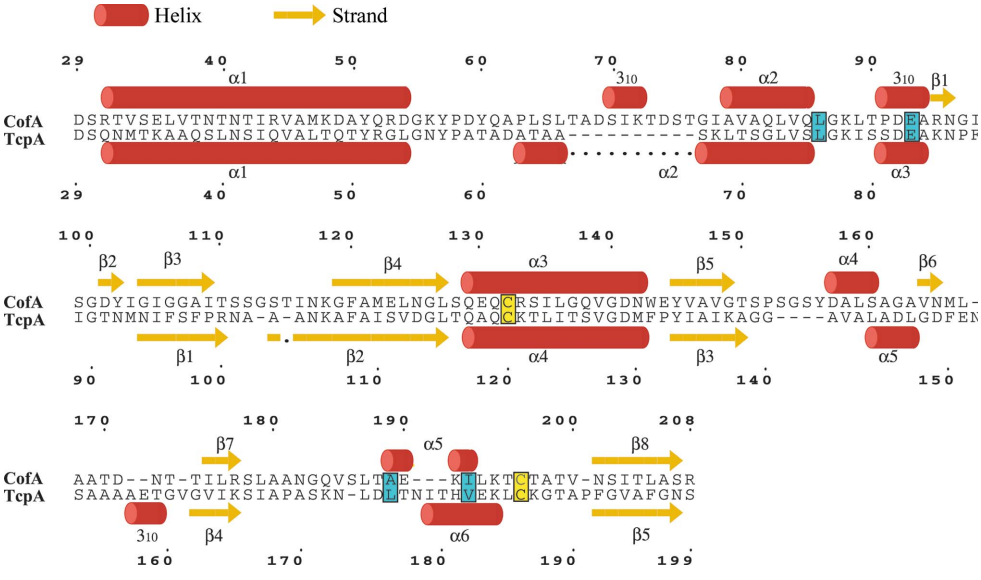


Figure 7
Sequence alignment between CofA and TcpA, which was generated by the *MATRAS* server (Kawabata, 2003) by applying pairwise three-dimensional alignment. *DSSP* was used for secondary-structural assignment (Kabsch & Sander, 1983). The important residues for pilus assembly suggested by a previous study of TcpA are highlighted with a cyan background. Cysteine residues making disulfide bonds are shown on a yellow background.

terminal α-helix and Glu93 in the $\alpha\beta$-loop (Fig. 6*d*). Although we only constrained the N-terminal residues during structural refinement, all of the key interactions involving residues Leu86 and Glu93 between the globular head domains that were previously reported in the TcpA filament model were also successfully formed in the CofA model. This result signifies the importance and accessibility of these residues in subunit–subunit recognition (Fig. 6).

Our model also newly identified tight hydrophobic interactions between Leu86 in the $\alpha\beta$-loop of one subunit and two residues (Ala189 and Ile192) in the α5 helix of the neighbouring subunit. Sequence alignment showed that the hydrophobic residues Ala189 and Ile192 were conserved as Leu176 and Val182, respectively, at the corresponding positions in TcpA (Fig. 7). A previous study showed that a mutation of Leu76 (corresponding to Leu86 in CofA) to Lys prevents TcpA from forming pili. These results indicate that the interface comprising the $\alpha\beta$-loop and D-region could be stabilized by both electrostatic and hydrophobic interactions, including those involving the abovementioned residues (Li *et al.*, 2008).

Despite recent progress in revealing the structure of type IV pili, the mechanism by which these filaments are made is still not fully understood. Once pilin subunits are synthesized in the cytosol, they should be delivered and transported across the inner membrane (mainly *via* Sec machinery) to participate in the growing filament (Allen *et al.*, 2012). During this pilus biogenesis, the N-terminal hydrophobic segment initially remains anchored in the membrane by hydrophobic interactions. Interactions involving the $\alpha\beta$-loop and D-region of the globular head domains, which possibly protrude towards the periplasm, may provide a driving force to initiate the addition of each subunit to the growing pilus (Li *et al.*, 2008). Our results observed during the structure-refinement process showed that the proposed interaction among globular head domains was successfully formed with no *a priori* structural constraint in this region if the N-terminal α-helices were tightly anchored. We speculated that if subunit–subunit interactions at the globular head domain were successfully formed then this condition might help to anchor the N-terminal α-helices and to promote pilus growth.

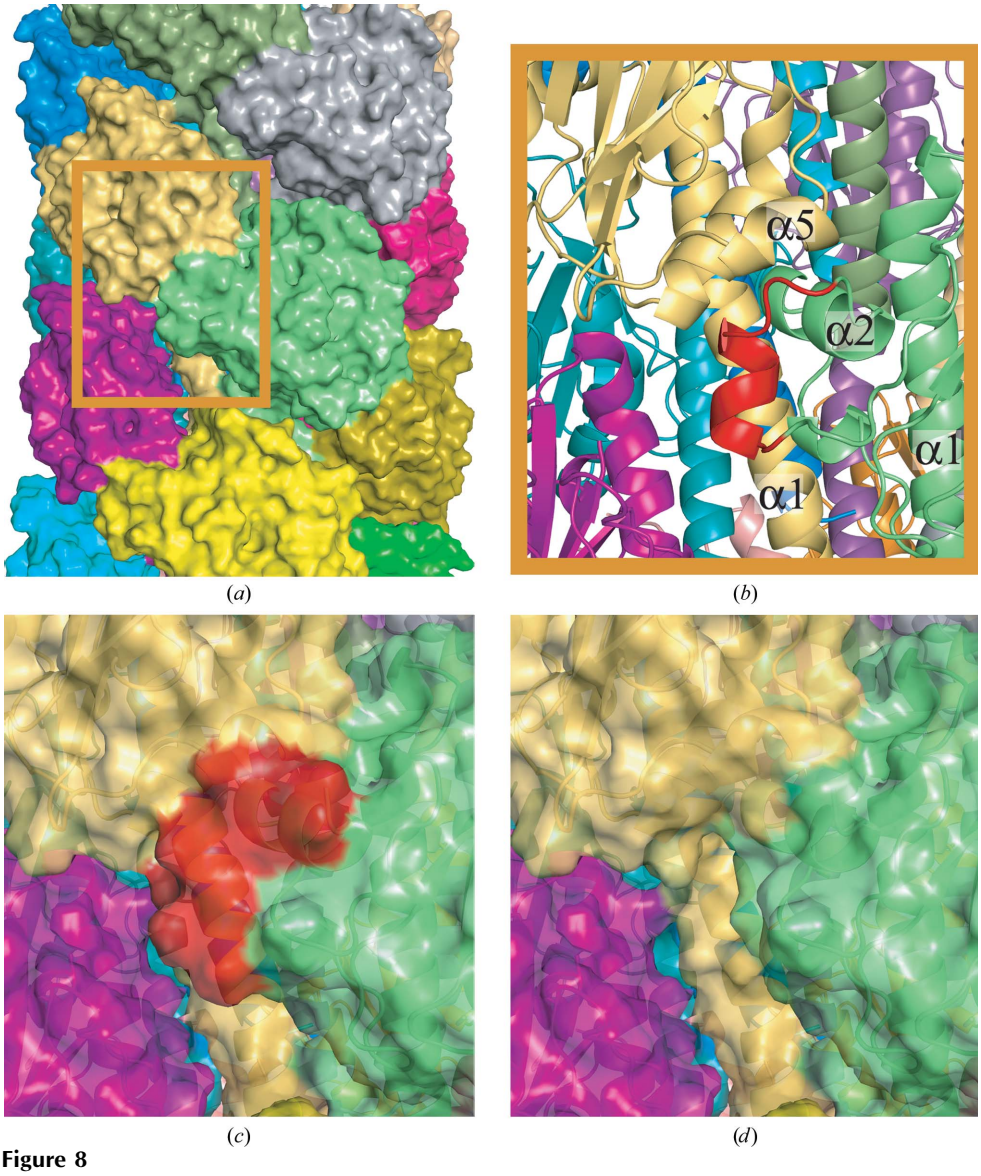


Figure 8
Surface of the CofA filament. (*a*) van der Waals (vdW) surface of the CofA model. (*b*) Enlarged ribbon model of the area surrounded by the orange box in the CofA surface model in (*a*). The insertion loop, which is found in the $\alpha\beta$-loop region of CofA, is shown in red. (*c*) Surface model of the enlarged area in (*b*). (*d*) A surface model without the insertion loop of CofA is shown to simulate the groove of the TcpA filament model.

3.5. Comparison of the pilus surface of CofA with that of TcpA

Two unique features of the CofA pilus filament were observed on its surface. The TcpA pilus filament displays a characteristic rectangular gap between the pilin subunits, in which the polar surface of an amphipathic segment of the N-terminal α-helix is solvent-accessible (Li *et al.*, 2008). This gap is predicted to be important for pilus–pilus interactions in TcpA. The protruding D-region of one filament presumably intercalates into the gap, interacting with the amphipathic segment of the N-terminal

α-helix. A van der Waals surface model of the CofA pilus (Fig. 8) showed that this gap was partially blocked by the protruding loop between the α1 and α2 helices, which was exceptionally inserted into CofA. This scenario prevented interaction between the protruding D-regions and N-terminal α-helix. Therefore, if a pilus–pilus interaction exists in CofA then it is potentially different from that of TcpA.

In addition, the surface electrostatic potential of CofA was largely different from that of TcpA, apparently reflecting their low sequence identity (~30%; Fig. 9). In fact, the surface of the CofA filament was highly negatively charged; the region that was positively charged was almost negligible. This feature is distinct from other type IVb pilins, such as TcpA, PilS and BfpA, in which the solvent-exposed surface is composed of both positively and negatively charged regions. Although the detailed function of the pilus filament of CofA is still largely uncharacterized, the observed differences at the molecular surface reflect the distinct adhesive function of CofA when compared with other type IVb pilin family members.

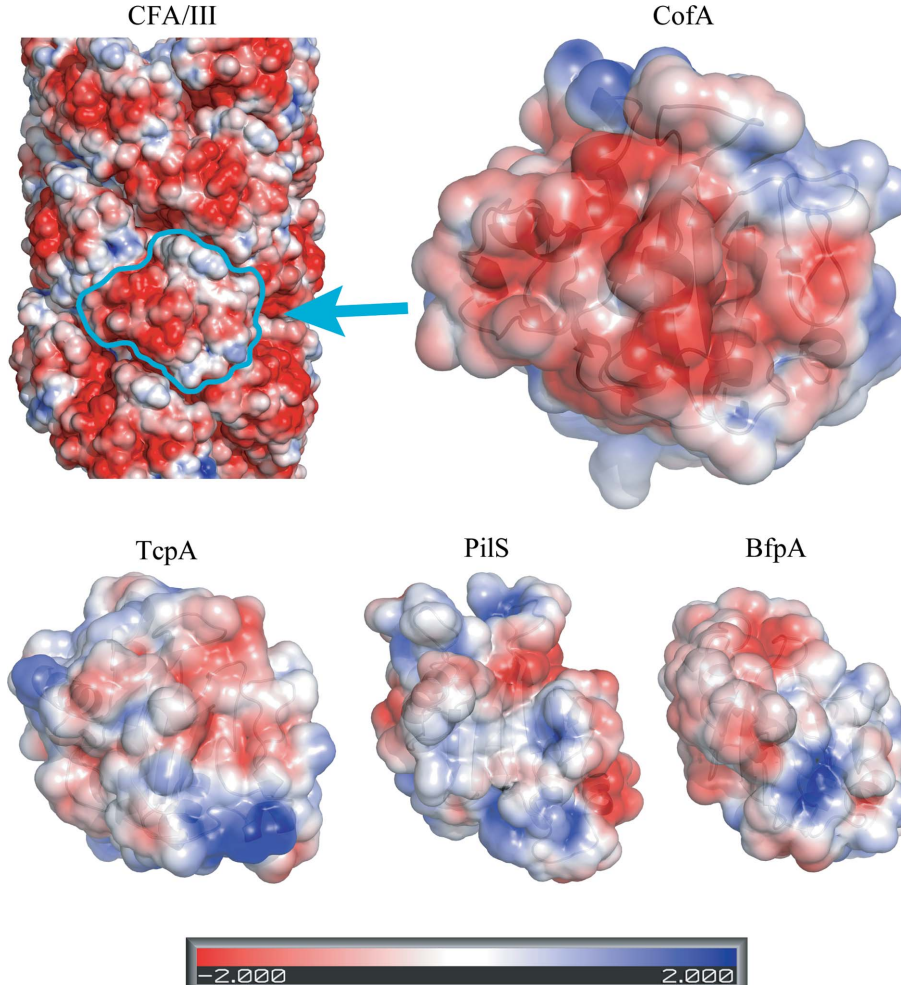


Figure 9
Electrostatic potentials of the CofA filament model and other type IVb pilins at the solvent-accessible surface, which were generated by the *Adaptive Poisson–Boltzmann Solver* (*APBS*; Baker *et al.*, 2001) plug-in in *PyMOL*. The orientation of each pilin subunit is similar to that in Fig. 4.

4. Conclusions

In the present study, we determined the crystal structure of the major pilin subunit CofA of CFA/III in ETEC by sulfur-SAD phasing at the extremely high resolution of 0.90 Å in space group $P2_1$. While this manuscript was under review, Kolappan and coworkers also reported a crystal structure of CofA which was determined by the molecular-replacement method using a number of modified model of TcpA at a resolution of 1.26 Å in space group $P1$ (Kolappan *et al.*, 2012). Despite the differences in crystal packing, these structures are essentially the same. Remarkably, several results, including these newly available CofA structures, indicate that the core structure of type IVb pilins is highly conserved, comprising a long N-terminal hydrophobic α-helix and an antiparallel β-sheet. However, owing to the marginal overall sequence homology among the type IVb pilins, CofA also exhibits structural variations. These variations are especially notable in the $\alpha\beta$-loop and D-region, both of which are solvent-exposed and important for pilus assembly. As exemplified by our CofA filament model, these structural differences produce unique surface properties for this filament compared with other type IVb pili. Thus, it is tempting to speculate that the functions and assembly mechanisms of pilin subunits could diverge depending on their bacterial origin.

The synchrotron X-ray experiments were performed with the approval of the SPring-8 Program Advisory Committee (2011B1259). This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) and by the Program of Founding Research Centres for Emerging and Re-emerging Infectious Diseases from MEXT.

References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.
Allen, W. J., Phan, G. & Waksman, G. (2012). *Curr. Opin. Struct. Biol.* **22**, 500–506.
Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
Balakrishna, A. M., Saxena, A. M., Mok, H. Y.-K. & Swaminathan, K. (2009). *Proteins*, **77**, 253–261.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). *Acta Cryst.* D**59**, 2023–2030.
Campos, M., Francetic, O. & Nilges, M. (2011). *J. Struct. Biol.* **173**, 436–444.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Cowtan, K. D. & Zhang, K. Y. J. (1999). *Prog. Biophys. Mol. Biol.* **72**, 245–270.
Craig, L., Pique, M. E. & Tainer, J. A. (2004). *Nature Rev. Microbiol.* **2**, 363–378.
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. & Tainer, J. A. (2003). *Mol. Cell*, **11**, 1139–1150.
Dhakal, B. K., Bower, J. M. & Mulvey, M. A. (2009). *Encyclopedia of Microbiology*, edited by M. Schaechter, pp. 470–489. Oxford: Elsevier.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Engh, R. A. & Huber, R. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 382–392. Dordrecht: Kluwer Academic Publishers.
Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M., Pieper, U. & Sali, A. (2006). *Curr. Protoc. Bioinformatics*, Unit 5.6.
Fleckenstein, J. M., Hardwidge, P. R., Munson, G. P., Rasko, D. A., Sommerfelt, H. & Steinsland, H. (2010). *Microbes Infect.* **12**, 89–98.
Gaastra, W. & Svennerholm, A. M. (1996). *Trends Microbiol.* **4**, 444–452.
Girón, J. A., Gómez-Duarte, O. G., Jarvis, K. G. & Kaper, J. B. (1997). *Gene*, **192**, 39–43.
Gómez-Duarte, O. G., Ruiz-Tagle, A., Gómez, D. C., Viboud, G. I., Jarvis, K. G., Kaper, J. B. & Girón, J. A. (1999). *Microbiology*, **145**, 1809–1816.
Hazes, B., Sastry, P. A., Hayakawa, K., Read, R. J. & Irvin, R. T. (2000). *J. Mol. Biol.* **299**, 1005–1017.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. (2008). *J. Chem. Theory Comput.* **4**, 435–447.
Higashi, D. L., Biais, N., Weyand, N. J., Agellon, A., Sisko, J. L., Brown, L. M. & So, M. (2011). *PLoS One*, **6**, e21373.
Jackson, R. M. & Sternberg, M. J. (1995). *J. Mol. Biol.* **250**, 258–275.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
Kawabata, T. (2003). *Nucleic Acids Res.* **31**, 3367–3369.
Kirn, T. J., Lafferty, M. J., Sandoe, C. M. & Taylor, R. K. (2000). *Mol. Microbiol.* **35**, 896–910.
Kolappan, S., Roos, J., Yuen, A. S., Pierce, O. M. & Craig, L. (2012). *J. Bacteriol.* **194**, 2725–2735.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
Li, J., Lim, M. S., Li, S., Brock, M., Pique, M. E., Woods, V. L. Jr & Craig, L. (2008). *Structure*, **16**, 137–148.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
Parge, H. E., Forest, K. T., Hickey, M. J., Christensen, D. A., Getzoff, E. D. & Tainer, J. A. (1995). *Nature (London)*, **378**, 32–38.
Pelicic, V. (2008). *Mol. Microbiol.* **68**, 827–837.
Qadri, F., Svennerholm, A. M., Faruque, A. S. & Sack, R. B. (2005). *Clin. Microbiol. Rev.* **18**, 465–483.
Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 1020–1027.
Ramboarina, S., Fernandes, P. J., Daniell, S., Islam, S., Simpson, P., Frankel, G., Booy, F., Donnenberg, M. S. & Matthews, S. (2005). *J. Biol. Chem.* **280**, 40252–40260.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.
Taniguchi, T., Akeda, Y., Haba, A., Yasuda, Y., Yamamoto, K., Honda, T. & Tochikubo, K. (2001). *Infect. Immun.* **69**, 5864–5873.
Taniguchi, T., Arita, M., Sato, M., Yamamoto, K., Miwatani, T. & Honda, T. (1994). *J. Infect. Dis.* **170**, 1049–1050.
Taniguchi, T., Fujino, Y., Yamamoto, K., Miwatani, T. & Honda, T. (1995). *Infect. Immun.* **63**, 724–728.
Touhami, A., Jericho, M. H., Boyd, J. M. & Beveridge, T. J. (2006). *J. Bacteriol.* **188**, 370–377.
Wang, J. W., Chen, J. R., Gu, Y. X., Zheng, C. D. & Fan, H. F. (2004). *Acta Cryst.* D**60**, 1991–1996.
Wang, J., Dauter, M. & Dauter, Z. (2006). *Acta Cryst.* D**62**, 1475–1483.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Xu, X.-F., Tan, Y.-W., Lam, L., Hackett, J., Zhang, M. & Mok, Y.-K. (2004). *J. Biol. Chem.* **279**, 31599–31605.