Received 24 November 2014

Accepted 24 March 2015

Edited by A. Nakagawa, Osaka University, Japan

Keywords: ETEC; type IVb pili; minor pilin; CFA/III; CofB.

Cloning, expression, purification, crystallization and X-ray crystallographic analysis of CofB, the minor pilin subunit of CFA/III from human enterotoxigenic *Escherichia coli*

Kazuki Kawahara,^a Hiroya Oki,^a Shunsuke Fukakusa,^{a,b} Takahiro Maruno,^c Yuji Kobayashi,^c Daisuke Motooka,^d Tooru Taniguchi,^d Takeshi Honda,^d Tetsuya Iida,^d Shota Nakamura^{d*} and Tadayasu Ohkubo^a

^aGraduate School of Pharmaceutical Sciences, Osaka University, 1-6 Yamadaoka, Suita, Osaka 565-0871, Japan, ^bCenter for Research of Ancient Culture, Nara Women's University, Kita-Uoya-Nishi Machi, Nara, Nara 630-8506, Japan, ^cGraduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan, and ^dResearch Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan. *Correspondence e-mail: nshotan@gen-info.osaka-u.ac.jp

Colonization factor antigen III (CFA/III) is one of the virulence factors of human enterotoxigenic *Escherichia coli* (ETEC) that forms the long, thin, proteinaceous fibres of type IV pili through assembly of its major and minor subunits CofA and CofB, respectively. The crystal structure of CofA has recently been reported; however, the lack of structural information for CofB, the largest among the known type IV pilin subunits, hampers a comprehensive understanding of CFA/III pili. In this study, constructs of wild-type CofB with an N-terminal truncation and the corresponding SeMet derivative were cloned, expressed, purified and crystallized. The crystals belonged to the rhombohedral space group *R*32, with unit-cell parameters $a = b = 103.97$, $c = 364.57$ Å for the wild-type construct and $a = b = 103.47$, $c = 362.08$ Å for the SeMet-derivatized form. Although the diffraction quality of these crystals was initially very poor, dehydration of the crystals substantially improved the resolution limit from ~ 4.0 to ~ 2.0 Å. The initial phase was solved by the single-wavelength anomalous dispersion (SAD) method using a dehydrated SeMet CofB crystal, which resulted in an interpretable electron-density map.

1. Introduction

Enterotoxigenic *Escherichia coli* (ETEC) has been widely recognized as a major bacterial cause of diarrhoea in children and travellers in developing countries (Qadri *et al.*, 2005). The pathogenicity of ETEC is attributable to its ability to adhere to and colonize the small intestine, where it produces enterotoxins such as heat-labile enterotoxin (LT) and/or heat-stable enterotoxin (ST) (Fleckenstein *et al.*, 2010). In human ETEC, these abilities depend on the presence of colonization-factor antigens (CFAs) and putative colonization factors (PCFs), which form pili (or fimbriae) to attach to the target cell surface (Gaastra & Svennerholm, 1996).

Previously, Taniguchi and coworkers determined the nucleotide sequence of the genetic region responsible for CFA/III formation of human ETEC and demonstrated that this *cof* operon consists of a cluster of 14 genes (Taniguchi *et al.*, 2001). Several of the determined protein sequences of the genes in the *cof* operon were found to be homologous to proteins related to type IV pilus biogenesis in other Gram-negative bacteria. These include toxin-coregulated pili (TCP) from *Vibrio cholerae*, bundle-forming pili (BFP) from enteropathogenic *E. coli* (EPEC) and the long pilus (longus)

© 2015 International Union of Crystallography

Table 1
Macromolecule-production information.

Source organism	Enterotoxigenic <i>E. coli</i> strain 31-10
Forward primer†	5'-CCCCCGGGGATGAAGCCAGACACAAA-TTG-3'
Reverse primer†	5'-CGCCGCTCGAGTTATTAGGTTTGTGGTTCTG-TACTGC-3'
Expression vector	pET-48b
Expression host	<i>E. coli</i> SHuffle (C3030)
Complete amino-acid sequence of the generated construct	GPDEARRQIVSNALISEIAGIVDFVAEEQITVIE-QGIEKEITNPLYEQSSGIPYINRTTNKDLNST-MSTNASEFINWAGTSTRIFTRKYCISTGTQ-GNYEFSKDYIPCEEPAILSNSDLKIDRIDFVA-TDNTVGSATIERVDFILTFDKSNANESFYFSNY-VSSLEKAAEQHSISFKDIYVVERNSSGAAGWR-LTTISGKPLTFSGLSKNIGSLDKTKNYGLRLS-IDPNLGKFLRADGRVGAADKLCWNIDNKMSPGCLAAADDSGNNLVLTKGKGAKSNEPGLCWDLNTG-TSKLCLTQIEGKDNNDKASLIKLDNDGNPA-TMLANILVEEKSMSTDKKELRTIPNTIYAAF-SNSNASDLVITNPGNYIGNVTSEKGRIELNVQ-DCPVSPDGKHLPRLSASIASIVADTKDSNGK-YQADFSSLAGNRNSGGQLGYLSGTAIQVNSQSG-SKWYITATMGVDFPLTNTTYVYLNPKFLSVNI-TTWCSTEPQT

† The SmaI and XhoI cleavage sites in the forward and reverse primer, respectively, are underlined.

of ETEC (Girón *et al.*, 1997; Gómez-Duarte *et al.*, 1999; Taniguchi *et al.*, 1994, 1995).

In Gram-negative bacteria, type IV pili mediate diverse biological functions, including adhesion, surface motility, microcolony formation, biofilm formation and DNA uptake (Dhakal *et al.*, 2009; Higashi *et al.*, 2011). Type IV pili are long, thin, proteinaceous fibrils with a typical size of approximately 60–90 Å in diameter and are synthesized through the assembly of characteristic fibre-forming proteins called type IV pilins. Further classification divides the type IV pilins into two subgroups, types IVa and IVb, based primarily on their N-terminal signal sequence length and the size of the mature protein (Pelicic, 2008). Previously, we identified the genes in the *cof* operon encoding the major and minor type IV pilin subunits CofA and CofB, respectively. However, the structural details and the mechanism of assembly of these subunits to form type IV pili are still largely unknown. The development of novel vaccines against such bacterial pathogens critically relies on this information.

The crystal structure of CofA has recently been determined and was shown to adopt an α/β -roll fold typical of type IV pilin

subunits. The CofA structure also featured an N-terminal α -helix cradled against a conserved β -sheet consisting of four antiparallel β -strands (Fukakusa *et al.*, 2012; Kolappan *et al.*, 2012). Analysis of the CofA structural data provided evidence to explain the distinctive adhesive properties of the filaments of CFA/III compared with other type IV pili, with the discovery of a highly negatively charged surface potential (Fukakusa *et al.*, 2012; Kolappan *et al.*, 2012). Although some structural characteristics of CFA/III are now being identified, little is known about the functional role of the minor pilin subunit CofB, which could provide functional variety in CFA/III pili. Therefore, it is crucial to comprehensively understand the colonization and adhesive behaviour of human ETEC. In this study, we cloned, expressed, purified and crystallized a construct and a selenomethionyl (SeMet) derivative of CofB, in which the conserved hydrophobic N-terminal segment (~28 amino acids) was truncated to improve protein solubility.

2. Materials and methods

2.1. Cloning, expression and purification of the N-terminally truncated CofB, Δ N28-CofB

As noted previously (Fukakusa *et al.*, 2012), the highly conserved hydrophobic N-terminal portion of type IV pilins often makes these proteins insoluble. On the basis of a sequence alignment with other type IV pilins (Fig. 1), an N-terminally truncated construct of CofB was designed in which the first 28 amino acids were eliminated during cloning to solubilize the resulting protein. The DNA sequence encoding the truncated CofB (here termed Δ N28-CofB) was PCR-amplified from plasmid pTT240, which contains the *cof* operon (GenBank ID AB049751), using the sequence-specific primer pair listed in Table 1. The amplified PCR product was digested with SmaI and XhoI and ligated into a pET-48b expression vector (Merck Biosciences, Germany), resulting in both thioredoxin and 6 $\times$ His tags at the N-terminus. *E. coli* SHuffle T7 Express LysY cells (New England Biolabs, Ipswich, Massachusetts, USA) were transformed with the resulting expression vector. The *E. coli* cells were cultured in Luria–Bertani (LB) medium in the presence of kanamycin (20 μ g ml⁻¹) and chloramphenicol (10 μ g ml⁻¹) for selection until the culture reached an optimal density at 660 nm and 310 K of ~0.60.

Figure 1
Amino-acid sequence alignment of the hydrophobic N-terminal segment of type IV pilins. The N-terminal sequence (52 amino acids) of CofB from human ETEC is aligned with that of its major pilin counterpart, CofA, and those of other type IVb pilins, including LngA from ETEC, TcpA from *V. cholerae* and BfpA from EPEC. The N-terminal sequence of PAK pilin (categorized as type IVa) is also 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 CofB, Δ N28-CofB.

The culture was induced with 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) and incubated for 18 h at 293 K. The cells were centrifuged at 6000g for 15 min at 277 K and the cell pellet was resuspended in lysis buffer consisting of 50 mM Tris-HCl pH 8.0, 1 M NaCl, 1% (v/v) Triton X-100. The suspension was sonicated ten times with a Branson Sonifier 150 (Branson Ultrasonics, Danbury, Connecticut, USA) for 1 min on ice. The lysate was clarified by centrifugation and loaded onto a 5 ml HiTrap Chelating HP nickel-chelating column (GE Healthcare Biosciences, Pittsburgh, Pennsylvania, USA) that had been pre-equilibrated with modified lysis buffer [50 mM Tris-HCl pH 8.0, 1 M NaCl, 0.1% (v/v) Triton X-100]. After washing the column with wash buffer [50 mM Tris-HCl pH 8.0, 1 M NaCl, 0.1% (v/v) Triton X-100, 10 mM imidazole], the recombinant protein was eluted with a linear gradient of 10–500 mM imidazole and dialyzed against buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl). To remove the thioredoxin and 6 $\times$ His tag, the dialyzed sample was cleaved by HRV 3C protease (Takara Bio, Japan) at 277 K according to the manufacturer's protocol, with a modification of the reaction time to 72 h owing to the relatively high expression level of the protein. The resulting protein was purified using a 5 ml HiTrap Q HP anion-exchange column (GE Healthcare Biosciences) that had been pre-equilibrated with buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl) and was

(a)

(b)

Figure 2
(a) Crystal of N-terminally truncated CofB. (b) Crystal of SeMet-derivatized N-terminally truncated CofB.

Table 2

Data collection and processing.

Values in parentheses are for the outer shell.

	SeMet- Δ N28-CofB	Δ N28-CofB
Diffraction source	BL38B1, SPring-8	BL38B1, SPring-8
Wavelength ($\text{\AA}$)	0.97887	1.0000
Temperature (K)	100	100
Detector	ADSC Q315	ADSC Q315
Crystal-to-detector distance (mm)	200	220
Rotation range per image ($^\circ$)	1	1
Total rotation range ($^\circ$)	360	120
Exposure time per image (s)	6	50
Space group	<i>R</i> 32	<i>R</i> 32
<i>a</i> , <i>b</i> , <i>c</i> ($\text{\AA}$)	103.47, 103.47, 362.08	103.97, 103.97, 364.57
α , β , γ ($^\circ$)	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Mosaicity ($^\circ$)	0.191	0.172
Resolution range ($\text{\AA}$)	19.84–2.10 (2.15–2.10)	50.00–1.88 (1.99–1.88)
Total No. of observations	1030692	433256
No. of unique reflections	90541	117860
Completeness (%)	99.7 (98.6)	98.7 (99.6)
Multiplicity	11.38 (11.24)	3.67 (3.74)
$\langle I/\sigma(I) \rangle$	27.11 (7.21)	14.94 (3.77)
$R_{\text{merge}}^\dagger$	0.068 (0.355)	0.057 (0.336)
$R_{\text{r.i.m.}}^\ddagger$	0.073 (0.381)	0.067 (0.394)
Wilson <i>B</i> factor§ ($\text{\AA}^2$)	23.660	23.231
$CC_{1/2}$	0.999 (0.981)	0.999 (0.845)

$^\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th measurement of the same reflection and $\langle I(hkl) \rangle$ is the mean observed intensity for the reflection. $^\ddagger R_{\text{r.i.m.}} = \sum_{hkl} [N(hkl)/[N(hkl) - 1]]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $N(hkl)$ is the multiplicity (the number of times that reflection *hkl* was measured). § Overall *B* factor from Wilson plot ($\text{\AA}^2$) calculated using *SCALA* (Evans, 2006).

eluted with a linear gradient of 0–2 M NaCl. The recombinant protein was further purified using a size-exclusion column (Superdex 200 26/60, GE Healthcare Biosciences) that had been pre-equilibrated with buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl). Finally, the purity of Δ N28-CofB was verified by SDS-PAGE (data not shown) and the protein solution was concentrated to $\sim 12 \text{ mg ml}^{-1}$ with a 30 kDa molecular-weight cutoff Amicon Ultra centrifugal filtration unit (Millipore, Billerica, Massachusetts, USA). The protein concentration was determined from the absorbance at 280 nm ($\epsilon = 57340 \text{ mol}^{-1} \text{ cm}^{-1}$).

The SeMet derivative of the N-terminally truncated CofB, SeMet- Δ N28-CofB, was prepared according to the previously reported method utilizing SeMet minimal medium (Doublié, 1997). The purification procedure of the SeMet-derivatized protein was the same as that used for the Δ N28-CofB construct.

2.2. Crystallization

Initial crystallization conditions were screened by the sitting-drop vapour-diffusion method at 277 and 293 K using commercially available screening kits, including Crystal Screen HT, Index HT, SaltRx, Crystal Screen Cryo and Crystal Screen Cryo 2 (Hampton Research, Aliso Viejo, California, USA), JBScreen Classic I + II and JBScreen Cryo (Jena BioScience, Germany), NR-LBD, Clear Strategy I and II, Heavy + Light and Pact HT (Molecular Dimensions, Altamonte Springs, Florida, USA), Wizard Screen 1 and 2 and

Wizard Cryo 1 and 2 (Emerald Bio, Bainbridge Island, Washington, USA). Crystallization drops were prepared by mixing 1 μ l protein solution and 1 μ l reservoir solution and were equilibrated against 80 μ l reservoir solution. Preliminary crystals of Δ N28-CofB grew using a reservoir solution consisting of 4.0 M sodium formate (Hampton Research Crystal Screen HT condition No. 33). After several rounds of optimization, crystals of approximately $0.4 \times 0.2 \times 0.1$ mm in size were obtained from a mixture of 3 μ l protein solution (5 mg ml^{-1}) and an equal volume of reservoir solution consisting of 2.5 M sodium formate, 0.1 M sodium acetate pH 5.0 at 293 K after two weeks (Fig. 2a).

SeMet- Δ N28-CofB was crystallized using the same conditions as had been optimized for Δ N28-CofB crystals. Similarly shaped crystals grew to approximately $0.3 \times 0.2 \times 0.1$ mm in size after two weeks (Fig. 2b).

2.3. Data collection and processing

To improve the diffraction quality of the crystals, both Δ N28-CofB and SeMet- Δ N28-CofB crystals obtained from the crystallization drop containing 2.5 M sodium formate were dehydrated before X-ray diffraction experiments by transferring the hanging-drop cover slip to a new well and further equilibrating against fresh reservoir solution containing increasing concentrations of a dehydrating agent, either 4.0 or 5.0 M sodium formate, for at least 3 d. The preliminary in-house X-ray diffraction experiments revealed that the dehydrating agent could also act as a cryoprotectant (Bujacz *et al.*, 2010). This enabled the crystals to be retrieved from the droplet using a LithoLoop (Molecular Dimensions) and flash-cooled by direct placement into a cold nitrogen-gas stream at 100 K.

For both the Δ N28-CofB and SeMet- Δ N28-CofB crystals, diffraction data were collected using an ADSC Quantum 315 CCD detector on beamline BL38B1 at SPring-8, Hyogo, Japan. All collected diffraction data were indexed and processed with *XDS* and were scaled with *XSCALE* (Kabsch, 1993). Data-collection and processing statistics are given in Table 2.

3. Results and discussion

The mature form of the minor pilin CofB has a sequence length of 518 amino acids, which is more than twice the length of the major pilin CofA (208 amino acids). To our knowledge, CofB (~ 50 kDa) is the largest type IV pilin reported to date (typical size of 7–20 kDa; Giltner *et al.*, 2012). Although homology searches using available databases show no structural homologues, the N-terminal hydrophobic segment (~ 50 amino acids) of CofB has meaningful sequence homology to those of other type IV pilins, including the glutamate at position 5 known to be critical for pilus assembly (Fig. 1). Nonetheless, the highly conserved N-terminal hydrophobic segment often makes these proteins insoluble and produces protein aggregates, hampering protein crystallization (Fukushima *et al.*, 2012). Therefore, in this study we designed the

construct Δ N28-CofB, in which the first 28 residues at the N-terminus of CofB were truncated.

The Δ N28-CofB and SeMet- Δ N28-CofB constructs were highly soluble in solution and were successfully crystallized in a trapezoid-like shape with a typical thickness of ~ 0.1 mm (Fig. 2). X-ray diffraction experiments for crystals of both constructs performed on beamline BL38B1 at SPring-8 initially showed very poor diffraction patterns, with a maximum resolution of up to 4.0 Å at a wavelength of 1.0 Å. Therefore, to improve the diffraction quality of the crystals, several post-crystallization treatments such as crystal annealing and dehydration were performed (Heras *et al.*, 2003, 2005). In particular, a dehydration procedure involving the one-step transfer of the cover slip with the crystal droplet to a new well, which was then equilibrated against fresh reservoir solution containing increasing concentration of the dehydrant, *i.e.* 5.0 M sodium formate, over the course of 3 d substantially improved the diffraction quality of the crystals. As an example, the highest resolution of the SeMet-derivative crystal was improved from ~ 3.5 to ~ 2.0 Å at a wavelength of 0.97887 Å using this protocol (Fig. 3).

The SeMet- Δ N28-CofB crystals belonged to the rhombohedral space group *R*32, with unit-cell parameters $a = 103.47$, $b = 103.47$, $c = 362.08$ Å. Assuming that the asymmetric unit contained one molecule, the Matthews coefficient was $3.51 \text{ Å}^3 \text{ Da}^{-1}$ and the solvent content of the crystal was estimated to be 65.0%. To alleviate radiation damage to the crystal, single-wavelength anomalous dispersion (SAD), instead of multi-wavelength anomalous dispersion (MAD),

Figure 3 X-ray diffraction pattern of an SeMet-derivative crystal of the N-terminally truncated CofB before the dehydration procedure; the crystal diffracted to a maximum resolution of ~ 3.5 Å at a wavelength of 0.97887 Å. The inset highlights a representative high-resolution diffraction area of the SeMet-derivative crystal before and after the dehydration procedure.

data were collected from an SeMet-derivative crystal. Initial diffraction images, collected at a wavelength of 0.97887 Å and a crystal-to-detector distance of 200 mm, indicated that the SeMet-derivative crystal diffracted well to a resolution of ~1.80 Å when a sufficiently long exposure time of 50 s was used. To maximize the data redundancy and reduce the possible radiation damage, however, the diffraction data for SAD phasing were collected with a relatively short exposure time of 6.0 s for each image. Owing to this limitation, the data were processed to a maximum resolution of 2.10 Å, which was slightly lower than the resolution limit expected for the SeMet-derivative crystal.

Initial phase calculations were performed using the programs *SHELXC*, *SHELXD* and *SHELXE* from the *SHELX* suite (Sheldrick *et al.*, 2008), which produced interpretable electron-density maps. Model-building and refinement of the structure using higher resolution diffraction data collected from the ΔN28-CofB crystal are now in progress.

Acknowledgements

We thank Dr Seiki Baba of the Japan Synchrotron Radiation Research Institute for his help during data collection. The synchrotron X-ray diffraction experiments were performed with the approval of the SPring-8 Proposal Review Committee (2012A1379, 2012B1217 and 2013A1251).

References

- Bujacz, G., Wrzesniewska, B. & Bujacz, A. (2010). *Acta Cryst.* **D66**, 789–796.
- Dhakal, B. K., Bower, J. M. & Mulvey, M. A. (2009). *Encyclopedia of Microbiology*, edited by M. Schaechter, pp. 470–489. Oxford: Elsevier.
- Doublé, S. (1997). *Methods Enzymol.* **276**, 523–530.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Fleckenstein, J. M., Hardwidge, P. R., Munson, G. P., Rasko, D. A., Sommerfelt, H. & Steinsland, H. (2010). *Microbes Infect.* **12**, 89–98.
- Fukakusa, S., Kawahara, K., Nakamura, S., Iwashita, T., Baba, S., Nishimura, M., Kobayashi, Y., Honda, T., Iida, T., Taniguchi, T. & Ohkubo, T. (2012). *Acta Cryst.* **D68**, 1418–1429.
- Gaastra, W. & Svennerholm, A. M. (1996). *Trends Microbiol.* **4**, 444–452.
- Giltner, C. L., Nguyen, Y. & Burrows, L. L. (2012). *Microbiol. Mol. Biol. Rev.* **76**, 740–772.
- 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.
- Heras, B., Edeling, M. A., Byriel, K. A., Jones, A., Raina, S. & Martin, J. L. (2003). *Structure*, **11**, 139–145.
- Heras, B. & Martin, J. L. (2005). *Acta Cryst.* **D61**, 1173–1180.
- Higashi, D. L., Biais, N., Weyand, N. J., Agellon, A., Sisko, J. L., Brown, L. M. & So, M. (2011). *PLoS One*, **6**, e21373.
- Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
- Kolappan, S., Roos, J., Yuen, A. S., Pierce, O. M. & Craig, L. (2012). *J. Bacteriol.* **194**, 2725–2735.
- Pellicic, V. (2008). *Mol. Microbiol.* **68**, 827–837.
- Qadri, F., Svennerholm, A. M., Faruque, A. S. W. & Sack, R. B. (2005). *Clin. Microbiol. Rev.* **18**, 465–483.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- 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.