Rui Bao,^a Lothar Esser,^a Steven Poole,^b Annette McVeigh,^b Yu-xing Chen,^c Stephen J. Savarino^{b,d} and Di Xia^{a*}

^aLaboratory of Cell Biology, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD 20892-4256, USA, ^bEnteric Diseases Department, Infectious Diseases Directorate, Naval Medical Research Center, Silver Spring, MD 20895-7500, USA, ^cHefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China, and ^dDepartment of Pediatrics, Uniformed Services, University of the Health Sciences, Bethesda, MD 20814-4799, USA

Correspondence e-mail: xiad@mail.nih.gov

Received 24 October 2013

Accepted 16 December 2013

Preliminary X-ray diffraction analysis of CfaA, a molecular chaperone essential for the assembly of CFA/I fimbriae of human enterotoxigenic *Escherichia coli*

Understanding of pilus bioassembly in Gram-negative bacteria stems mainly from studies of P pili and type 1 fimbriae of uropathogenic *Escherichia coli*, which are mediated by the classic chaperone–usher pathway (CUP). However, CFA/I fimbriae, a class 5 fimbria and intestinal colonization factor for enterotoxigenic *E. coli* (ETEC), are proposed to assemble *via* the alternate chaperone pathway (ACP). Both CUP and ACP fimbrial bioassembly pathways require the function of a periplasmic chaperone, but their corresponding proteins share very low similarity in primary sequence. Here, the crystallization of the CFA/I periplasmic chaperone CfaA by the hanging-drop vapor-diffusion method is reported. X-ray diffraction data sets were collected from a native CfaA crystal to 2 Å resolution and to 1.8 and 2.8 Å resolution, respectively, from a lead and a platinum derivative. These crystals displayed the symmetry of space group C2, with unit-cell parameters $a = 103.6$, $b = 28.68$, $c = 90.60$ Å, $\beta = 119.7^\circ$. Initial phases were derived from multiple isomorphous replacement with anomalous scattering experiments using the data from the platinum and lead derivatives. This resulted in an interpretable electron-density map showing one CfaA molecule in an asymmetric unit. Sequence assignments were aided by anomalous signals from the heavy-atom derivatives. Refinement of the atomic model of CfaA is ongoing, which is expected to further understanding of the essential aspects and allowable variations in tertiary structure of the greater family of chaperones involved in chaperone–usher mediated bioassembly.

1. Introduction

The chaperone–usher pathway (CUP) is an evolutionarily conserved mechanism for surface fimbria and/or fibril biogenesis of many Gram-negative bacteria (Nuccio & Bäumlér, 2007). It requires a minimum of three components: a structural pilin subunit that is capable of polymerizing into bacterial surface fimbriae/fibrils, a periplasmic chaperone that stabilizes pilin subunits and escorts them to the outer membrane for assembly, and an outer membrane usher that orchestrates ordered tip-to-base assembly (Choudhury *et al.*, 1999; Bao *et al.*, 2013; Kuehn *et al.*, 1993). Bacterial fimbriae or pili, also referred to as colonization factors (CFs; Evans *et al.*, 1975), are essential virulence factors for pathogenesis, and to date over 22 different CFs have been identified for enterotoxigenic *Escherichia coli* (ETEC; Gaastra & Svennerholm, 1996; Isidean *et al.*, 2011). The colonization factor antigen I (CFA/I) fimbriae of ETEC, a major cause of diarrhea in travelers to and in young children in resource-limited countries (Huilan *et al.*, 1991; Wennerås & Erling, 2004), are the most prevalent surface-adhesive organelles among isolated ETEC field strains (Black, 1990). CFA/I fimbria is an archetype of a family (class 5) of eight related but serologically distinct fimbriae (Anantha *et al.*, 2004) that share similar genetic and biochemical features and use a proposed mechanism dubbed the ‘alternate chaperone pathway’ (ACP) for assembly (Sakellaris & Scott, 1998; Sakellaris *et al.*, 1996; Soto & Hultgren, 1999). Included in the class 5 fimbriae are coli surface antigen 1 (CS1), CS2, CS4, CS14, CS17, CS19 and the putative CF O71 (PCFO71). CFA/I biogenesis requires four protein components, a major (CfaB) and a minor (CfaE) structural subunit and two assembly proteins: the chaperone (CfaA) and the usher (CfaC).

The bioassembly of CFA/I fimbriae is initiated by the attachment of the chaperone/minor pilin (CfaA/CfaE) heterodimer to the usher protein CfaC at the outer membrane assembly site followed by cyclic incorporation of the major pilin (CfaB) (Poole *et al.*, 2007; Li *et al.*, 2007, 2009). Both pilin subunits, CfaE and CfaB, require the chaperone CfaA for folding and stability and are incorporated into the pilus by the donor-strand exchange mechanism (Sauer *et al.*, 1999; Choudhury *et al.*, 1999; Kuehn *et al.*, 1993). The ability of chaperones to provide stability to individual pilin subunits is by means of donor-strand complementation, which was revealed by crystallographic analyses of chaperone–pilin complexes from P pili and type 1 fimbriae of uropathogenic *E. coli* and from fraction 1 (F1) antigen of *Yersinia pestis*, leading to the classification of FGS and FGL chaperones (Chapman *et al.*, 1999). In spite of these advances, the use of ACP for biogenesis of the class 5 fimbriae implies structural and functional distinctions for chaperone proteins of the class 5 fimbriae, and to date there is no structural information available for any of these chaperones. Here, we report the crystallization and preliminary X-ray diffraction analysis for the chaperone CfaA of CFA/I fimbriae. The detailed structural knowledge of the chaperone–subunit interaction is expected to further our understanding of the assembly process of CFA/I fimbriae, which is expected to impact the development of potential therapeutic interferences.

2. Materials and methods

2.1. Cloning, expression and purification of recombinant CfaA

The expression plasmid pET24a-*cfaA*(*his*)₆ was constructed by incorporating the full-length *cfaA* coding region (Gene ID 12796579) derived from wild-type plasmid pNTP513 (Hibberd *et al.*, 1991), adding an LEHHHHHH purification tag C-terminally and transforming into *E. coli* strain BL21(DE3) (New England Biolabs, Ipswich, Massachusetts, USA). The resultant *E. coli* cells were grown in Terrific broth (Research Products International, Mount Prospect, Illinois, USA) in the presence of 50 µg ml⁻¹ kanamycin at 310 K to an optical density of 1.5 at 600 nm, at which point the culture was cooled

to 291 K and CfaA expression was induced by adding IPTG to a final concentration of 0.8 mM. After a further 16 h of incubation at 291 K, cells were collected by centrifugation at 4000g for 25 min. The cell pellets were resuspended in a hypertonic buffer consisting of 60 mM Tris–HCl pH 7.5, 20% glycerol for 10 min followed by another centrifugation. The periplasmic extracts were prepared by resuspending the cell pellet in an ice-cold hypotonic buffer consisting of 5 mM MgCl₂, 5 mM CaCl₂, 20 mM Tris–HCl pH 7.5, 50 mM NaCl followed by high-speed centrifugation at 16 000g for 30 min. The supernatant was loaded onto a Ni–NTA Superflow column (Qiagen, Germantown, Maryland, USA) pre-equilibrated with binding buffer (25 mM Tris–HCl pH 7.5, 100 mM NaCl) plus 20 mM imidazole. His-tagged CfaA was eluted with the same binding buffer plus 300 mM imidazole. As a last step, size-exclusion chromatography was used to further purify CfaA and the resulting CfaA was concentrated to 10 mg ml⁻¹ with an Amicon Ultra-15 centrifugal filtration device with 10 kDa molecular-weight cutoff (Millipore, Billerica, Massachusetts, USA).

2.2. Crystallization, preparation of heavy-metal derivatives, X-ray diffraction image collection and data reduction

Crystallization of purified CfaA was initiated with a robotic screen of 576 crystallization conditions using the vapor-diffusion hanging-drop method in 96-well plate format (Greiner Bio-One, Monroe, North Carolina, USA), in which 150 nl commercial crystallization solution (Crystal Screen HT, Index HT, PEG/Ion HT, SaltRx HT, Hampton Research, Aliso Viejo, California, USA; Structure Screen 1 and 2 HT, Molecular Dimensions, Altamonte Springs, Florida, USA; Wizard I and II, Emerald BioSystems, Bedford, Massachusetts, USA) were admixed with 150 nl protein solution (10 mg ml⁻¹) using a robotic liquid dispenser (Mosquito, TTP LabTech, England) and incubated at 293 K over 60 µl reservoir solution. Optimization of initial hits was carried out manually with Linbro plates (Hampton Research, Aliso Viejo, California, USA) by mixing 2–4 µl protein solution with 2 µl optimized well solution. Crystals usually appeared after 5 d of incubation at 293 K. Heavy-metal derivatives were prepared by soaking native crystals in well solution plus 2 mM K₂PtCl₄ or 15 mM lead(II) acetate overnight. Native and derivatized crystals were stabilized by cross-linking using 0.15% glutaraldehyde solution in a vapor-diffusion setup (Lusty, 1999). Cross-linked crystals were further treated with a buffer containing 30% glycerol prior to flash-cooling in liquid propane, which allows a faster rate of cooling (Teng & Moffat, 1998).

Diffraction images were collected on a MAR CCD 225 detector at the SER-CAT beamline 22D of the Advanced Photon Source (APS), Argonne National Laboratory (ANL). Data reduction including indexing, integration and scaling was performed using the *HKL-2000* package (Otwinowski & Minor, 1997). The initial phase was solved experimentally by multiple isomorphous replacement coupled with anomalous scattering (MIRAS) using *SOLVE/RESOLVE* as implemented in the *PHENIX* suite (Terwilliger, 2003; Adams *et al.*, 2010).

3. Results and discussion

3.1. Protein purification, crystallization, cryoprotection and preliminary X-ray diffraction analysis

The chaperone protein CfaA of the human-specific ETEC CFA/I fimbriae was cloned and its C-terminus was extended with a hexahistidine tag *via* a dipeptide linker (LE), resulting in the expression plasmid pET24a-*cfaA*(*his*)₆. The protein was overexpressed in *E. coli* and was then purified to homogeneity by nickel-affinity chromato-

Figure 1
SDS–PAGE of purified recombinant CfaA. The right lane contains molecular-mass marker (labelled in kDa).

graphy followed by size-exclusion chromatography. The protein yield was about 8 mg per litre of *E. coli* culture. The purity of the protein preparation was checked by SDS-PAGE (Fig. 1). Both N-terminal sequencing and mass-spectrometric analysis of the purified protein confirmed the cleavage of the signal peptide. Mature CfaA with a C-terminal hexahistidine tag has 227 residues with a molecular weight of 26 kDa. Purified CfaA was concentrated to approximately 10 mg ml^{-1} , with which crystallization screens were performed robotically using commercial crystallization screening kits at 291 K. Three conditions yielded promising hits (A11 and F10 from Hampton Research Index HT and A8 from Emerald Wizard I and II HT). CfaA tends to form crystal clusters (Fig. 2a). Optimization of initial hits led to the final conditions (22% PEG 3350, 200 mM NaCl, 100 mM MES pH 5.3) that allowed reproducible crystallization of single crystals of better morphology growing from large crystal clusters (Fig. 2b).

Flash-cooling of CfaA crystals with liquid nitrogen led to an increase in the crystal mosaic spread and poor diffraction, which prompted a search for cryoprotection conditions that could minimize the increase in mosaic spread, improve diffraction quality and

enhance compatibility with the subsequent heavy-metal-derivatization process. We found that cross-linking with glutaraldehyde was particularly helpful to provide mechanical strength to CfaA crystals for cryoprotection. After cross-linking, CfaA crystals were treated with a cryoprotectant consisting of 20% PEG 3350, 0.15 M NaCl, 70 mM MES pH 5.3, 30% glycerol before flash-cooling in liquid propane.

Crystals of CfaA typically diffracted X-rays to $1.8 \text{ \AA}$ resolution (Fig. 3) using synchrotron radiation. The diffraction patterns are consistent with the symmetry of a *C*-centered monoclinic crystal lattice (space group *C2*). The native crystals had unit-cell parameters $a = 103.6$, $b = 28.68$, $c = 90.60 \text{ \AA}$, $\beta = 119.7^\circ$. Calculation of the Matthews coefficient (V_M) suggested the presence of one ($V_M = 2.25 \text{ \AA}^3 \text{ Da}^{-1}$) CfaA molecule per asymmetric unit (Matthews, 1968).

3.2. Search for heavy-atom-derivatized CfaA crystals

Attempts to solve the crystallographic phase problem by the molecular-replacement method failed to yield a correct solution owing to a lack of suitable search models. Thus, other experimental phasing approaches must be adopted. The presence of a single methionine residue (Met3) at the N-terminus of the mature CfaA sequence precludes the use of straightforward SeMet derivatization for phasing. Our first approach was to dope heavy-metal compounds directly into CfaA crystals. A large number of compounds such as $\text{Hg}(\text{CH}_3\text{COO})_2$, HgCl_2 , $\text{K}_2\text{Pt}(\text{NO}_2)_4$, K_2HgI_4 , AgNO_3 , K_2PtCl_4 , $\text{Pb}(\text{CH}_3\text{COO})_2$, K_2PtCl_6 and $(\text{CH}_3)_3\text{Pb}(\text{CH}_3\text{COO})$ were tested and most of them led to a decline in diffraction quality.

In a separate approach, attempts were made to introduce methionine mutations into the CfaA, in which we aligned eight sequences of chaperone proteins from the class 5 family fimbriae and constructed two mutants based on the principle of conservative substitution of hydrophobic residues. One mutant contained four substitutions (I4M, L213M, I214M and L216M) and another had two mutations (I51M and L190M), all based on the numbering of the

Figure 2
(a) Clusters of CfaA crystals from initial crystallization screening. (b) CfaA crystals after optimization.

Figure 3
Diffraction image from an optimized CfaA crystal. The diffraction resolution at the edge of the image is about $1.8 \text{ \AA}$.

Table 1

Statistics for X-ray diffraction data sets and phasing.

The space group of the CfaA crystals is *C2*. All diffraction data sets were collected at beamline 22-BM, Argonne National Laboratory. Values in parentheses are for the outermost shells.

Data sets	Native	Pb	Pt
Wavelength (Å)	0.98197	0.98197	0.89197
Resolution (Å)	50–2.03 (2.08–2.03)	30–1.77 (1.81–1.77)	50–2.80 (2.90–2.80)
Unit-cell parameters (Å, °)	<i>a</i> = 103.60, <i>b</i> = 28.68, <i>c</i> = 90.60, β = 119.70	<i>a</i> = 102.12, <i>b</i> = 28.51, <i>c</i> = 90.04, β = 118.95	<i>a</i> = 102.04, <i>b</i> = 28.64, <i>c</i> = 90.41, β = 119.01
No. of observations	51359	138094	105508
No. of unique reflections	14999 (732)	22616 (1450)	5596 (202)
Average mosaicity (°)	0.22	0.20	0.20
$R_{\text{merge}}^{\dagger}$	0.066 (0.433)	0.053 (0.394)	0.084 (0.205)
Completeness (%)	97.4 (72.9)	99.3 (97.9)	95.1 (69.4)
Average $I/\sigma(I)$	19.3 (2.7)	19.6 (2.2)	24.5 (4.6)
Multiplicity	3.4 (2.9)	3.3 (2.7)	6.4 (3.4)

$\dagger R_{\text{merge}}$ is defined as $\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th observation of a reflection with Miller indices *hkl* and $\langle I(hkl) \rangle$ is the mean intensity for all measured *I(hkl)* and Friedel pairs.

Table 2

MIRAS phasing statistics.

	Sites	<i>x</i>	<i>y</i>	<i>z</i>	Occupancy	<i>B</i> (Å ²)	PP _{centric} [†]	PP _{acentric}
Pt derivative	1	0.3984	0.3933	0.4867	0.7739	59.1	0.83	0.76
Pb derivative	1	0.9701	0.3316	0.3011	0.0761	15.9	0.38	0.41
	2	0.4647	0.1313	0.0253	0.2622	14.9		
Resolution (Å)		10.22	6.41	4.99	4.23	3.73	3.10	2.89
Mean figure of merit		0.61	0.64	0.58	0.54	0.50	0.44	0.40

$\dagger$ Phasing power (PP) is defined as $\langle f_H \rangle / \text{residual}$, where $\langle f_H \rangle$ is the root-mean-square (r.m.s.) of the structure-factor contribution to amplitudes from heavy atoms in the derivative and residual is the r.m.s. of the lack-of-closure error.

mature protein sequence. *E. coli* cells bearing mutant constructs grew slowly and had considerably lower expression levels of CfaA.

In the meantime, the search for stabilization conditions led to a buffer solution (22% PEG 3350, 0.2 *M* NaCl, 0.1 *M* MES pH 5.3) in which CfaA crystals appeared to be stable. Heavy-atom soaking into stabilized CfaA crystals was repeated at different heavy-metal concentrations and for various lengths of time. These crystals were then cross-linked and cryoprotected. Diffraction data sets were collected and analyzed for potential derivatives. Two data sets turned out to be isomorphous derivatives to the native data set: one from overnight soaking with 2 mM K₂PtCl₄ and the other from soaking with 15 mM Pb(CH₃COO)₂.

3.3. Phase determination

A native crystal diffracted X-rays to better than 2 Å resolution, and the resulting data set was processed to 2.07 Å resolution with an R_{merge} of 6.6% (Table 1). As many as 12 diffraction data sets were collected from various heavy-metal-compound-soaked crystals, but only two were useful for phasing (Tables 1 and 2). One data set was obtained from a lead-derivatized crystal at 1.77 Å resolution and a second data set was obtained from a crystal treated with a platinum compound to 2.8 Å resolution (Table 1 and 2). The similarities in unit-cell parameters between derivative and native crystals (Table 1) and the small isomorphous *R* factors (R_{iso}) of 0.24 and 0.30 for the platinum and lead derivatives, respectively, indicated that both derivatives were relatively isomorphous to the native data set. Both derivative data sets and the native data set were used for MIRAS

phasing. Two and one metal-binding sites were identified with reasonable occupancies for the Pb and Pt derivatives, respectively, giving rise to an overall figure of merit of 0.49 in the resolution range between 10.22 and 2.89 Å. Phase improvement by density modification and phase extension was subsequently carried out, producing a high-quality electron-density map with clearly recognizable features of a protein molecule suitable for automated chain tracing. The Pb and Pt data sets were also used to compute anomalous difference Fourier maps with phase information from the improved density-modified MIRAS phases. Residues with heavy-atom attachment were unambiguously identified and were helpful in sequence assignment in model building, which is currently in progress. Based on the experimental electron-density map, there is one molecule in the asymmetric unit, consistent with the prediction by Matthews cell-content analysis.

The authors wish to thank the beamline staff of the SER-CAT at APS, ANL for their assistance in data collection. This research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research, by a grant from the Trans-NIH/FDA Intramural Biodefense Program (to DX), by the US Army Military Infectious Diseases Research Program Work Unit Number A0307 (to SJS) and by the Henry M. Jackson Foundation for the Advancement of Military Medicine (SJS).

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Anantha, R. P., McVeigh, A. L., Lee, L. H., Agnew, M. K., Cassels, F. J., Scott, D. A., Whittam, T. S. & Savarino, S. J. (2004). *Infect. Immun.* **72**, 7190–7201.
- Bao, R., Nair, M. K. M., Tang, W.-K., Esser, L., Sadhukhan, A., Holland, R. L., Xia, D. & Schifferli, D. M. (2013). *Proc. Natl Acad. Sci. USA*, **110**, 1065–1070.
- Black, R. E. (1990). *Rev. Infect. Dis.* **12**, S73–S79.
- Chapman, D. A., Zavialov, A. V., Chernovskaya, T. V., Karlyshev, A. V., Zav'yalova, G. A., Vasiliev, A. M., Dudich, I. V., Abramov, V. M., Zav'yalov, V. P. & MacIntyre, S. (1999). *J. Bacteriol.* **181**, 2422–2429.
- Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S. J. & Knight, S. D. (1999). *Science*, **285**, 1061–1066.
- Evans, D. G., Silver, R. P., Evans, D. J. Jr, Chase, D. G. & Gorbach, S. L. (1975). *Infect. Immun.* **12**, 656–667.
- Gaastra, W. & Svennerholm, A. M. (1996). *Trends Microbiol.* **4**, 444–452.
- Hibberd, M. L., McConnell, M. M., Willshaw, G. A., Smith, H. R. & Rowe, B. (1991). *J. Gen. Microbiol.* **137**, 1963–1970.
- Huilan, S., Zhen, L. G., Mathan, M. M., Mathew, M. M., Olarte, J., Espejo, R., Khin Maung, U., Ghafoor, M. A., Khan, M. A., Sami, Z. & Sutton, R. G. (1991). *Bull. World Health Organ.* **69**, 549–555.
- Isidean, S. D., Riddle, M. S., Savarino, S. J. & Porter, C. K. (2011). *Vaccine*, **29**, 6167–6178.
- Kuehn, M. J., Ogg, D. J., Kihlberg, J., Slonim, L. N., Flemmer, K., Bergfors, T. & Hultgren, S. J. (1993). *Science*, **262**, 1234–1241.
- Li, Y.-F., Poole, S., Nishio, K., Jang, K., Rasulova, F., McVeigh, A., Savarino, S. J., Xia, D. & Bullitt, E. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 10793–10798.
- Li, Y.-F., Poole, S., Rasulova, F., McVeigh, A. L., Savarino, S. J. & Xia, D. (2007). *J. Biol. Chem.* **282**, 23970–23980.
- Lusty, C. J. (1999). *J. Appl. Cryst.* **32**, 106–112.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Nuccio, S. P. & Bäuml, A. J. (2007). *Microbiol. Mol. Biol. Rev.* **71**, 551–575.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Poole, S. T., McVeigh, A. L., Anantha, R. P., Lee, L. H., Akay, Y. M., Pontzer, E. A., Scott, D. A., Bullitt, E. & Savarino, S. J. (2007). *Mol. Microbiol.* **63**, 1372–1384.
- Sakellaris, H., Balding, D. P. & Scott, J. R. (1996). *Mol. Microbiol.* **21**, 529–541.
- Sakellaris, H. & Scott, J. R. (1998). *Mol. Microbiol.* **30**, 681–687.
- Sauer, F. G., Fütterer, K., Pinkner, J. S., Dodson, K. W., Hultgren, S. J. & Waksman, G. (1999). *Science*, **285**, 1058–1061.
- Soto, G. E. & Hultgren, S. J. (1999). *J. Bacteriol.* **181**, 1059–1071.
- Teng, T.-Y. & Moffat, K. (1998). *J. Appl. Cryst.* **31**, 252–257.
- Terwilliger, T. C. (2003). *Methods Enzymol.* **374**, 22–37.
- Wennerås, C. & Erling, V. (2004). *J. Health Popul. Nutr.* **22**, 370–382.