STRUCTURAL BIOLOGY

ISSN 2059-7983

Structures of two fimbrial adhesins, AtfE and UcaD, from the uropathogen *Proteus mirabilis*

Wangshu Jiang,[a] Wimal Ubhayasekera,[a] Melanie M. Pearson[b] and Stefan D. Knight[a]*

[a]Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, PO Box 596, SE-751 24 Uppsala, Sweden, and [b]Department of Microbiology and Immunology, University of Michigan Medical School, 5641 Medical Science Building II, 1150 West Medical Center Drive, Ann Arbor, MI 48109-0620, USA. *Correspondence e-mail: stefan.knight@icm.uu.se

Received 14 July 2018
Accepted 3 September 2018

Edited by B. Kobe, University of Queensland, Australia

Keywords: fimbriae; adhesins; urinary-tract infection; *Proteus mirabilis*.

PDB references: $AtfE_{210}$, 6h1q; $UcaD_{211}$, 6h1x; $UcaD_{217}$, 6h2l

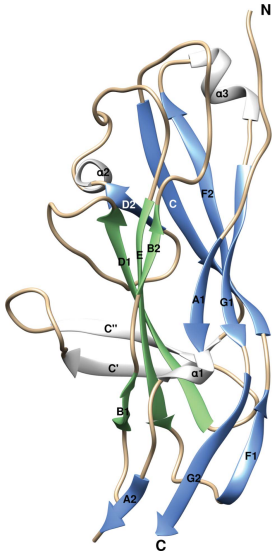
Supporting information: this article has supporting information at journals.iucr.org/d

The important uropathogen *Proteus mirabilis* encodes a record number of chaperone/usher-pathway adhesive fimbriae. Such fimbriae, which are used for adhesion to cell surfaces/tissues and for biofilm formation, are typically important virulence factors in bacterial pathogenesis. Here, the structures of the receptor-binding domains of the tip-located two-domain adhesins UcaD (1.5 Å resolution) and AtfE (1.58 Å resolution) from two *P. mirabilis* fimbriae (UCA/NAF and ATF) are presented. The structures of UcaD and AtfE are both similar to the F17G type of tip-located fimbrial receptor-binding domains, and the structures are very similar despite having only limited sequence similarity. These structures represent an important step towards a molecular-level understanding of *P. mirabilis* fimbrial adhesins and their roles in the complex pathogenesis of urinary-tract infections.

1. Introduction

The urinary tract is a common target for bacterial infections. Urinary-tract infections (UTIs) are clinically categorized as uncomplicated or complicated. Uncomplicated UTIs occur in individuals that are otherwise healthy, with uropathogenic *Escherichia coli* (UPEC) being the main causative agent. Complicated UTIs affect patients with underlying difficulties such as indwelling catheters. *Proteus mirabilis*, a Gram-negative bacterium famous for its ability to swarm across agar surfaces, is a major player in causing catheter-associated UTIs (CAUTIs; Nielubowicz & Mobley, 2010) and complicated UTIs, including pyelonephritis, bacteraemia and urolithiasis (Rocha *et al.*, 2007). *P. mirabilis* utilizes a multitude of virulence factors, including urease, flagella and fimbriae, to establish and promote infection (Schaffer & Pearson, 2015). Fimbriae are important virulence factors that are used for adhesion to cell surfaces/tissues and for biofilm formation (Schaffer & Pearson, 2015). 17 fimbrial gene clusters (FGCs) encoding chaperone/usher (CU) fimbrial assembly machineries and associated fimbrial subunits have been identified in the genome sequence of *P. mirabilis* HI4320, a representative strain isolated from the urine of an elderly, long-term (≥30 d) catheterized patient (Pearson *et al.*, 2008).

The CU pathway is commonly used by pathogenic Gram-negative bacteria to assemble Ig-like protein subunits ('pilins') into fimbriae (Geibel & Waksman, 2014; Zav'yalov *et al.*, 2010). In CU-mediated assembly, a periplasmic chaperone assists the folding of the incomplete immunoglobulin (Ig)-like pilin subunits and delivers them in a high-energy intermediate form (Zavialov *et al.*, 2003) to the outer membrane usher,



where assembly and secretion to the cell surface takes place. CU fimbriae can be classified into two main groups. Fimbrial monoadhesins display a specialized two-domain adhesin (TDA) subunit at the tip of rod-like fimbriae. TDAs consist of an N-terminal domain (NTD) that is responsible for high-affinity binding to receptors (usually carbohydrates), joined through a short linker to a C-terminal pilin domain. Fimbrial polyadhesins are instead typically thin and flexible structures in which each fimbrial subunit can bind with weak affinity to one or two distinct host cell receptors. Multivalent binding of polyadhesins to their receptors provides tight binding *via* a considerable avidity effect. 14 of the 17 *P. mirabilis* FGCs encode a putative TDA and hence are likely to produce fimbrial monoadhesins if expressed. Two of the most studied of these are uroepithelial cell adhesin (UCA), also known as non-agglutinating fimbriae (NAF), and ambient temperature fimbriae (ATF).

UCA/NAF was first identified in 1986 (Wray *et al.*, 1986). The major adherence component was purified from the outer membrane fraction of a *P. mirabilis* preparation incubated with shed human uroepithelial cells. The isolated 17.5 kDa protein subunit was found to self-assemble as long and flexible rods of 4–6 nm in width. The *uca* FGC, which is homologous to the FGCs encoding the F17 family of fimbriae in UPEC (Cook *et al.*, 1995), encodes a major fimbrial subunit (UcaA), a chaperone (UcaB), an usher (UcaC), a TDA (UcaD) and a transcriptional regulator (UcaJ). An isogenic *ucaA* mutant was significantly impaired both in adherence to exfoliated human uroepithelial cells as well as in kidney colonization (Pellegrino *et al.*, 2013). UCA fimbriae were also found to contribute to fitness in a CAUTI mouse model (Armbruster *et al.*, 2017). Taken together, these findings suggest that UCA fimbriae are important for the establishment and progression of *P. mirabilis* UTI.

ATF are expressed from the *atf* FGC, which contains six genes, *atfABCDEJ*, encoding a major structural subunit (AtfA), a chaperone (AtfB), an usher (AtfC), a minor subunit (AtfD), a TDA (AtfE) and a transcriptional regulator (AtfJ) regulating ATF expression (Armbruster *et al.*, 2018). An *atfABC* mutant was compared with the wild type in both independent-challenge and co-challenge experiments in an ascending UTI mouse model, and no obvious difference was found in bacterial counts in either bladder or kidneys 7 d post-inoculation (Zunino *et al.*, 2000). This, together with the fact that ATF are maximally expressed in static culture at 23°C, suggests that rather than contributing to virulence, ATF play a role in environments external to a mammalian host (Zunino *et al.*, 2000). However, other studies have pointed to links between the production of ATF and MR/P fimbriae, which are a major contributor to bladder infection (Armbruster *et al.*, 2018), suggesting that ATF somehow also contribute to UTI. For instance, when ATF was discovered in 1994, AtfA was found to co-purify with MR/P fimbriae even when cells were grown in conditions favouring the expression of ATF (Massad *et al.*, 1994). An isogenic *mrpA* mutant unable to express MR/P fimbriae was required in order to isolate ATF (Massad *et al.*, 1994). A later study compared bladder colonization in *P. mirabilis* mutant strains engineered with constitutive or no MR/P fimbrial production (locked ON or OFF; Jansen *et al.*, 2004). Strikingly, AtfA was found on the surface of bacteria colonizing mouse bladders 2 d after inoculation with the MR/P OFF strain, despite the unfavourable expression temperature. The MR/P OFF strain adhered to lamina propria underlying exfoliated uroepithelium, a pattern that was drastically different compared with MR/P ON and wild-type *P. mirabilis*, which both preferentially target intact uroepithelium. Also, the *atfABC* mutant was found to be the most affected among a collection of isogenic fimbrial mutant strains (including wild type, *mrpA*, *pmfA* and *ucaA*) in terms of biofilm formation and the ability to swarm across latex catheter segments (Scavone *et al.*, 2016). Therefore, the contribution of ATF to *P. mirabilis* pathogenesis might have been underestimated.

Here, we present the first two atomic structures of TDA receptor-binding domains from *P. mirabilis* fimbriae: UcaD and AtfE. Our work provides an important stepping stone towards a molecular-level understanding of *P. mirabilis* fimbrial adhesins and their roles in the complex pathogenesis of this important uropathogen.

2. Materials and methods

2.1. Construct design and cloning

Since pilin domains are very unstable and tend to aggregate in the absence of a chaperone, a convenient way to obtain material for structural and functional studies of TDAs is to express the receptor-binding NTDs by themselves (see, for example, Buts *et al.*, 2003; Bouckaert *et al.*, 2005). Constructs encoding the NTDs of the AtfE and UcaD TDAs ($AtfE_{ntd}$ and $UcaD_{ntd}$, respectively) were designed with the N-terminal signal peptides intact to ensure correct maturation following translocation to the periplasm. Initial putative domain borders were assigned based on multiple sequence alignment with pilin domains in combination with threading using *Phyre*2 (Kelley *et al.*, 2015). In order to optimize the constructs for protein expression, constructs with different numbers of residues included at the C-terminus were screened for soluble expression at the Protein Science Facility (PSF), Karolinska Institute, Stockholm, Sweden (http://ki.se/en/mbb/protein-science-facility) using a ligation-free cloning methodology (Aslanidis & de Jong, 1990). DNA encoding target constructs was PCR-amplified from *P. mirabilis* HI4320 genomic DNA. T4 DNA polymerase was used to produce complementary ssDNA overhangs on the PCR product and plasmid (pNIC-CH2). The expression plasmid introduces an additional alanine between the target DNA insertion site and the C-terminal His_6 tag encoded by the plasmid. Sequence-verified constructs were tested for small-scale expression using the *E. coli* BL21(DE3) R3 pRARE2 phage-resistant strain. The results of the PSF construct screening are shown in Table 1. Protein from two PSF constructs for UcaD (psfUCAD-c002 and psfUCAD-c004) was used for crystallization. Since the AtfE PSF constructs provided only low-yield large-scale expression, the DNA encoding AtfE residues 1–210 and the C-terminal His_6 tag was PCR-amplified from

Table 1
Constructs and Protein Science Facility screening results.

FGC	TDA gene	Construct	End residue	Soluble small-scale expression†	Comments
fim4/uca/naf	PMI0533	psfUCAD-c001	Asn208	1	
		psfUCAD-c002	Arg211	3	Purified protein: UcaD(21–211)-AHHHHHH
		psfUCAD-c003	Val214	3	
		psfUCAD-c004	Gln217	3	Purified protein: UcaD(21–217)-AHHHHHH
fim13/atf	PMI2732	psfATFE-c001	Asn205	2	
		psfATFE-c002	Thr208	2	
		psfATFE-c003	Thr210	3	Re-cloned into the pEXP5-CT/TOPO (Invitrogen) vector and transformed into *E. coli* SHuffle Express to produce AtfE(25–210)-AHHHHHH
		psfATFE-c004	Arg214	3	

† Amounts of both totally expressed and purified protein were analysed by SDS–PAGE. Based on the SDS–PAGE results, the soluble expression levels were scored as 4, dominating band of target protein; 3, band much stronger than background; 2, band equal to/stronger than background; 1, band weaker than background; 0, no detected protein.

the corresponding PSF plasmid (psfATFE-c003; Table 1) and recloned into the pEXP5-CT/TOPO vector (Invitrogen, USA) for verification of the sequence. The verified plasmid was transformed into SHuffle Express T7 competent *E. coli* (K12 strain; New England BioLabs, USA) for expression, since the SHuffle strain provided higher expression levels in small-scale tests (10 ml LB culture) compared with BL21 strains. For simplicity, all constructs used for crystallization will be designated by the C-terminal end residue in the following.

2.2. Protein production and purification

For $UcaD_{ntd}$, two constructs were used for structure determination. The $UcaD_{211}$ and $UcaD_{217}$ constructs were expressed in the *E. coli* BL21(DE3) R3 pRARE2 phage-resistant strain by inoculating fresh lysogeny broth (LB) containing 50 µg ml^{-1} kanamycin with a 1% volume of overnight pre-culture. The cells were incubated at 37°C with shaking until the OD_{600} reached 0.9. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 m*M*. Induced expression was carried out at 18°C overnight.

For $AtfE_{ntd}$, the $AtfE_{210}$ construct was expressed in SHuffle Express T7 competent *E. coli* cells (New England Biolabs, USA) by inoculating fresh LB medium containing 100 µg ml^{-1} ampicillin with a 1% volume of overnight pre-culture. The cells were incubated at 30°C (the recommended temperature for expression in SHuffle Express cells) with shaking until the OD_{600} reached 1.2. The cells were then transferred to 4°C for 20 min to cool before IPTG was added to a final concentration of 0.5 m*M* and expression continued at 18°C overnight.

The cells were harvested by centrifugation at 7500*g* at 4°C for 15 min. The harvested cells were homogenized in solubilization buffer [100 m*M* NaCl, 10 m*M* 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5, 5% glycerol for $UcaD_{ntd}$ and 500 m*M* NaCl, 50 m*M* HEPES pH 7.8, 5 m*M* dithiothreitol (DTT) for $AtfE_{ntd}$], to which one tablet of cOmplete protease-inhibitor cocktail (Roche) and 10 µg ml^{-1} DNase (from bovine pancreas; Sigma–Aldrich) were added. The cell suspension was stirred at 4°C for 45 min before passage through a cell disruptor (Constant Systems, UK) twice at 207 MPa pressure to ensure adequate lysis. The cell lysate was centrifuged at 30 500*g* for 30 min at 4°C and the supernatant was collected.

Soluble fractions of both $UcaD_{ntd}$ and $AtfE_{ntd}$ were first purified by nickel-immobilized metal-affinity chromatography (Ni-IMAC). The imidazole concentration in the soluble fraction was adjusted to 5 and 20 m*M*, respectively, before binding to equilibrated Ni Sepharose 6 Fast Flow (GE Healthcare) resin. Impurities and target proteins were separated by adding increasing concentrations of imidazole to the buffer. Unbound protein and most impurities were washed away with 50 m*M* imidazole. $UcaD_{ntd}$ and $AtfE_{ntd}$ were eluted with 300 and 500 m*M* imidazole, respectively. $UcaD_{ntd}$ was further purified using a Superdex 75 16/60 column (GE Healthcare) equilibrated with the homogenization buffer.

Ni-IMAC-eluted $AtfE_{ntd}$ was buffer-exchanged into 2-morpholinoethanesulfonic acid (MES) buffer (50 m*M* MES pH 5.6, 10 m*M* β-mercaptoethanol). A precipitate mainly consisting of impurities caused by the shift from pH 7.8 to 5.6 was removed by centrifugation at 6500*g* using a desktop centrifuge (Heraeus Multifuge 3SR with swing-out rotor 75006441). The protein was further purified using a HiTrap Heparin column (GE Healthcare), a two-mode chromatography agent with both affinity and cation-exchange capacity. $AtfE_{ntd}$ was eluted with a gradient of NaCl in 50 m*M* MES pH 5.6, 10 m*M* β-mercaptoethanol. Typically, $AtfE_{ntd}$ eluted at 600 m*M* NaCl.

2.3. Crystallization

Crystallization was carried out by sitting-drop vapour diffusion using a Mosquito Crystal crystallization robot (TTP Labtech, England). A typical crystallization drop was prepared by mixing protein and reservoir solutions in a 1:1 ratio to give a final volume of 300 nl. Seed stocks were prepared by diluting a drop containing crystalline material 660-fold by volume with the corresponding reservoir. Seeding was performed on the dispensed plate by adding 30 nl seed stock to each 300 nl drop.

Diffraction-quality $UcaD_{211}$ crystals appeared overnight at 293 K after the seeding of drops with protein at 20 mg ml^{-1} mixed with 0.005 *M* $CoCl_2$, 1 *M* HEPES pH 7.5, 10–15% PEG 3350. Initial crystalline material grown in 0.1 *M* trisodium

Table 2
Native data-collection and refinement statistics.

	$AtfE_{210}$	$UcaD_{211}$	$UcaD_{217}$
Data collection			
Space group	*C*2	$I4_122$	*P*1
a, *b*, *c* (Å)	74.91, 103.79, 67.76	73.35, 73.35, 153.11	30.47, 49.83, 61.83
α, β, γ (°)	90, 95.8, 90	90, 90, 90	86.45, 74.51, 75.52
Molecules in asymmetric unit	2	1	2
Wavelength (Å)	0.968	0.873	0.873
Resolution (Å)	33.7–1.58 (1.637–1.58)	41.89–1.70 (1.76–1.70)	37.40–1.50 (1.554–1.50)
Total reflections	502379 (50277)	170475 (16225)	192740 (18129)
Unique reflections	68833 (6417)	23405 (2269)	52581 (5178)
Multiplicity	7.3 (7.4)	7.3 (7.2)	3.7 (3.5)
Completeness (%)	97.25 (91.39)	99.75 (99.21)	96.61 (94.79)
$\langle I/\sigma(I)\rangle$	9.31 (0.82)	10.18 (1.20)	7.08 (1.30)
Wilson *B* factor (Å^2)	17.96	22.31	11.97
R_{merge}†	0.134 (2.10)	0.1162 (1.164)	0.1181 (0.9651)
R_{meas}‡	0.144 (2.25)	0.1251 (1.255)	0.1387 (1.14)
$R_{p.i.m.}$§	0.05293 (0.816)	0.04577 (0.4643)	0.07172 (0.5974)
$CC_{1/2}$	0.998 (0.500)	0.998 (0.699)	0.994 (0.530)
Refinement			
No. of reflections in work set	65012 (6079)	22133 (2149)	49977 (4900)
No. of reflections in test set	3409 (319)	1232 (112)	2570 (266)
R¶ (%)	17.8 (35.0)	19.8 (34.7)	20.9 (32.5)
R_{free}¶ (%)	20.2 (37.4)	22.0 (30.6)	23.8 (34.3)
No. of non-H atoms			
Total	3276	1671	3423
Protein	2861	1514	3062
Solvent	381	155	316
Ligands	4 glycerol, 2 PO_4^{3-}	2 Co^{2+}	9 SO_4^{2-}
R.m.s. deviations			
Bonds (Å)	0.007	0.005	0.007
Angles (°)	1.29	1.09	1.10
Ramachandran statistics			
Favoured (%)	98.29	98.42	97.90
Allowed (%)	1.71	1.58	2.10
Outliers (%)	0.00	0.00	0.00
Rotamer outliers (%)	1.26	0.00	0.30
Clashscore	2.79	3.30	1.77
Average *B* factor (Å^2)			
Overall	27.1	30.81	16.31
Protein	25.8	30.26	15.30
Ligands	48.8	47.04	42.94
Solvent	35.0	35.94	22.27
PDB code	6h1q	6h1x	6h2l

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. § $R_{p.i.m.} = \sum_{hkl}\{1/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ¶ $R = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$ calculated for a final model refined against all of the data. R_{free} is calculated for a test set comprising 5% of the reflections.

citrate pH 5.5, 20% PEG 3000 was used as a seed stock. The presence of cobalt was required for crystal formation. An iodine derivative of $UcaD_{211}$ for single anomalous dispersion (SAD) phasing was prepared by soaking the crystals in cryoprotectant containing NaI (0.005 *M* $CoCl_2$, 0.1 *M* HEPES pH 7.5, 15% PEG 3350, 20% glycerol, 1 *M* NaI) for 45 min before flash-cooling in liquid nitrogen.

$UcaD_{217}$ crystals for native data collection were obtained using protein at 16 mg ml^{-1} mixed with 2 *M* ammonium sulfate, 0.1 *M* bis-Tris pH 5.5. Diffraction-quality crystals appeared in a week at 293 K after seeding with seed stock from similar conditions. Crystals were cryoprotected with sucrose (by dropping a solid grain of sucrose into the crystallization drop) before flash-cooling with liquid nitrogen.

$AtfE_{210}$ crystals for native data collection were obtained using protein at 5 mg ml^{-1} in 0.8 *M* NaH_2PO_4, 0.8 *M* KH_2PO_4, 0.1 *M* HEPES pH 7.5. Diffraction-quality crystals appeared at 293 K 2–4 weeks after seeding with initial crystalline material obtained from similar conditions with 9 mg ml^{-1} $AtfE_{210}$. The crystals were cryoprotected with 18% glycerol before flash-cooling with liquid nitrogen. Crystals for iodine SAD phasing were grown in 0.7 *M* NaH_2PO_4, 0.7 *M* KH_2PO_4, 0.1 *M* sodium acetate pH 5.0 and were then soaked in 1 *M* NaI, 0.65 *M* NaH_2PO_4, 0.65 *M* KH_2PO_4, 0.1 *M* sodium acetate pH 5.0, 23% glycerol for 7 h prior to flash-cooling in liquid nitrogen.

2.4. Data collection and structure determination

Data collection was performed at 100 K at the European Synchrotron Radiation Facility (ESRF), Grenoble, France or at Diamond Light Source (DLS), UK. Diffraction data were processed using *XDS* (Kabsch, 2010) and scaled and merged using the *CCP*4 (Winn *et al.*, 2011) program *AIMLESS* (Evans & Murshudov, 2013). Experimental phasing and initial automatic model building was performed with *PHENIX* (Adams *et al.*, 2010) using the *phenix.autosol* (Terwilliger *et al.*, 2009) and *phenix.autobuild* (Terwilliger *et al.*, 2008) modules. *Phaser* (McCoy *et al.*, 2007) was used for molecular replacement. Refinement was carried out using *phenix.refine* (Afonine *et al.*, 2012). Torsional NCS restraints were applied during the refinement of structures with more than one protomer in the asymmetric unit ($UcaD_{217}$ and $AtfE_{210}$). All structures were modelled with isotropic temperature factors and TLS. Model building was performed using *Coot* (Emsley *et al.*, 2010). Structure validation was carried out using the tools available in *phenix.refine* (Afonine *et al.*, 2012) and *Coot* (Emsley *et al.*, 2010). Resolution cutoffs were based on $CC_{1/2}$ (Karplus & Diederichs, 2012) and the completeness of the data. Native data-collection and refinement statistics are listed in Table 2.

$UcaD_{211}$ crystallized in two crystal forms. Iodine SAD data for $UcaD_{211}$ were collected to 2.5 Å resolution on beamline I02 at DLS from a single crystal in space group $P4_322$ with two molecules in the asymmetric unit. Native $UcaD_{211}$ data were collected to 1.7 Å resolution on beamline ID23-2 at the ESRF from a single crystal in space group $I4_122$ with one molecule in the asymmetric unit. Native $UcaD_{217}$ data were collected to 1.5 Å resolution on beamline ID23-2 at the ESRF from a single crystal in space group *P*1 with two molecules in the asymmetric unit. The initial SAD-phased $UcaD_{211}$ model was

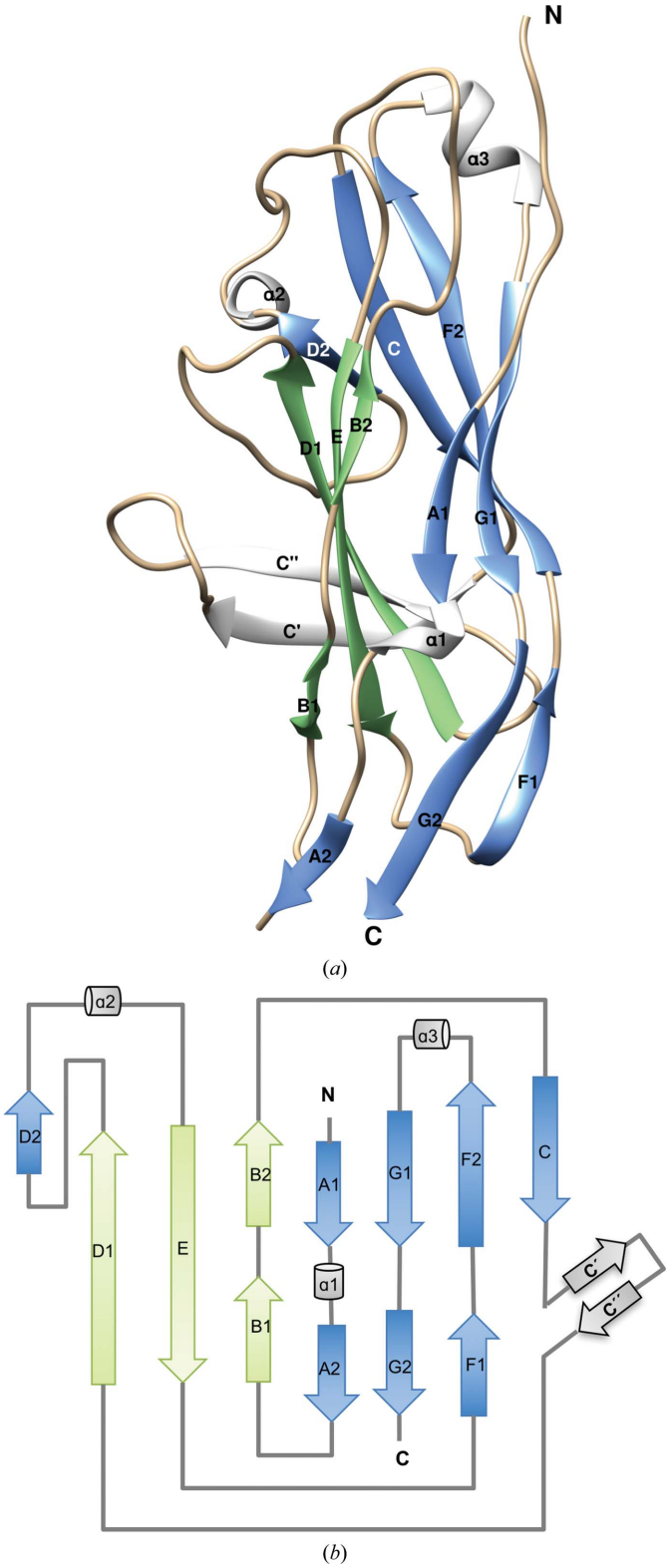


Figure 1
Three-dimensional crystal structure of $UcaD_{ntd}$. (*a*) Cartoon representation of $UcaD_{ntd}$. Secondary-structure elements are labelled A–G from the N-terminus to the C-terminus. β-Strands forming a β-sheet are in the same colour: β-strands A1/2, G1/2, F1/2, C and D2 are coloured blue, B1/2, E and D1 are coloured green and C′ and C″ are coloured grey. Helices are coloured grey. Coils are coloured pale yellow. (*b*) Topology diagram of $UcaD_{ntd}$ with the secondary-structure elements coloured in the same way as in (*a*).

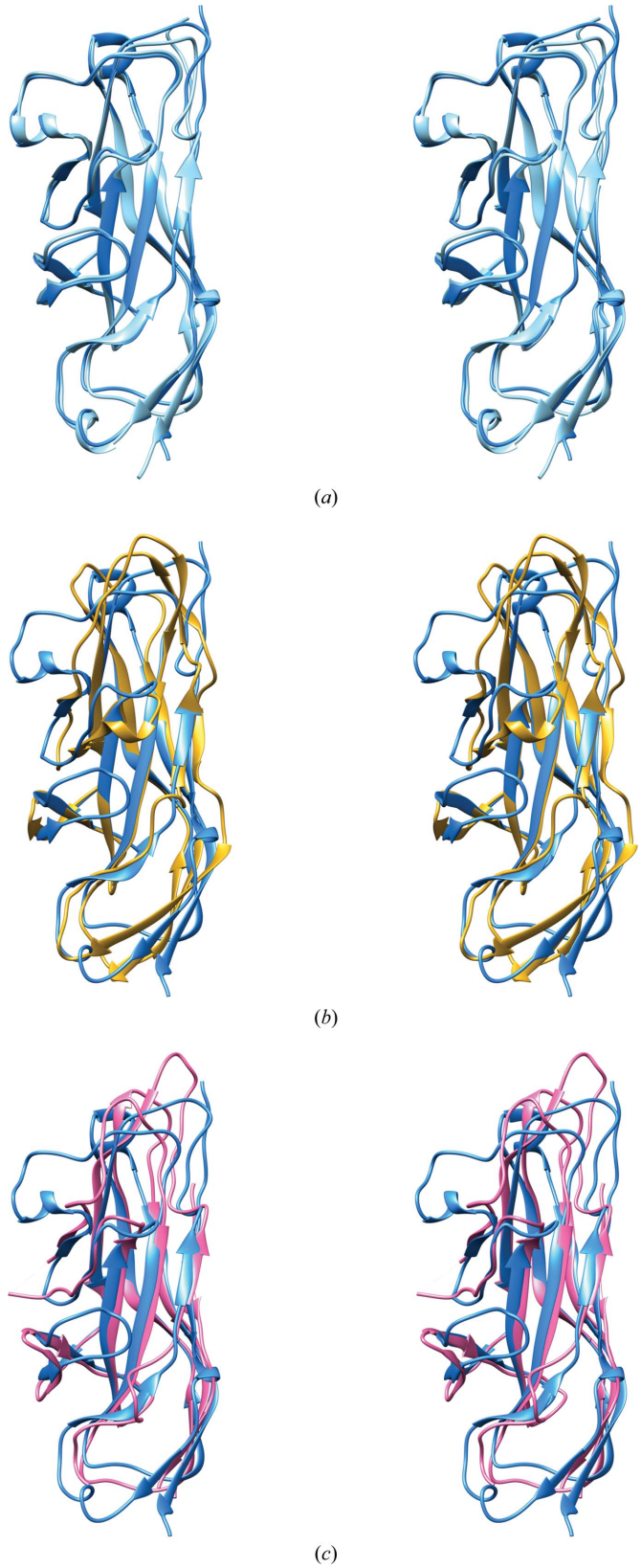


Figure 2
Structural superposition of the NTDs of (*a*) UcaD and UclD, (*b*) UcaD and F17cG, and (*c*) UcaD and AtfE shown as stereoviews. All structures are shown in carton representation and are superimposed using $UcaD_{ntd}$ as the reference. Colouring is based on sequence similarity. UcaD and UclD are shown in deep and light blue, respectively, F17cG is shown in gold and AtfE is shown in pink.

iteratively remodelled and refined against the native $UcaD_{211}$ data. The resulting model was used as a search model to solve the $UcaD_{217}$ structure by molecular replacement.

$AtfE_{210}$ crystallized in one crystal form belonging to space group *C*2 with two molecules in the asymmetric unit. Native $AtfE_{210}$ data were collected to 1.6 Å resolution from a single crystal on beamline ID30A-3 at the ESRF. Iodine SAD data were collected to 2.5 Å resolution from four $AtfE_{210}$ derivative crystals on beamline ID23-1 at the ESRF. The initial SAD-phased model was iteratively remodelled and refined against the native data.

2.5. Other software

Structures were superimposed and structure-based sequence alignments were obtained using *Chimera MatchMaker* (Meng *et al.*, 2006). Multiple sequence alignment was performed with *COBALT* (Papadopoulos & Agarwala, 2007) and was visualized using *ESPript* 3.0 (Robert & Gouet, 2014). Graphical representations of protein structures were prepared using *UCSF Chimera* (Pettersen *et al.*, 2004).

3. Results and discussion

Both $UcaD_{ntd}$ and $AtfE_{ntd}$ were expressed with their native N-terminal signal peptides to ensure correct maturation following translocation to the periplasm. In all cases the first residue visible in the electron-density maps corresponded to the first residue after the predicted signal sequence. The start and end points of the crystallized proteins are listed in Table 1.

3.1. Structure of $UcaD_{ntd}$

Since no structural model with sufficient sequence identity was available in the Protein Data Bank (PDB) at the time, *de novo* SAD phasing using a sodium iodide derivative of $UcaD_{211}$ that diffracted to 2.5 Å resolution was used to obtain initial phases. Similar to other TDA receptor-binding domains, $UcaD_{211}$ adopts an Ig-like Greek-key fold. However, the

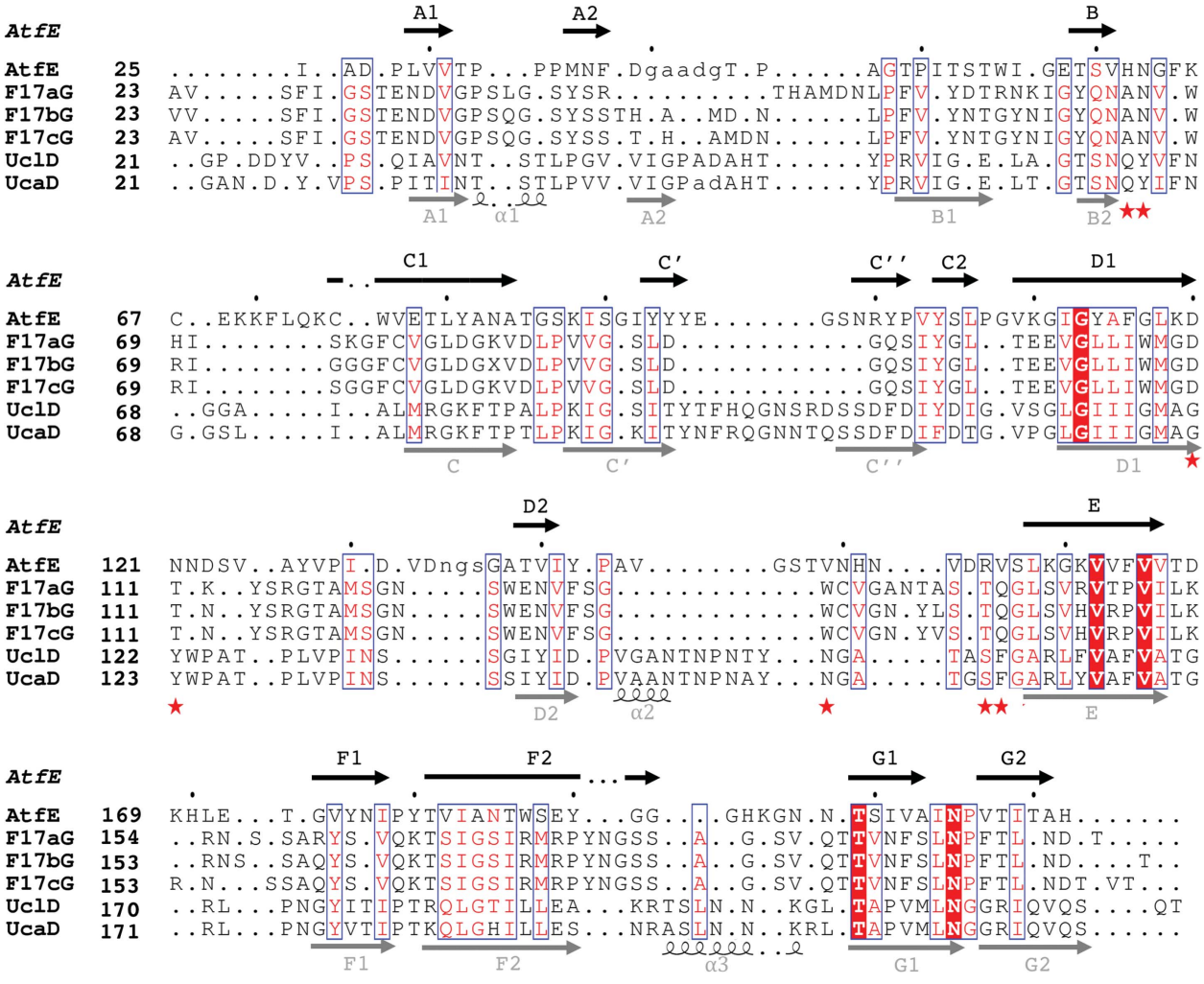


Figure 3
Structure-based multiple sequence alignment of the NTDs of AtfE, F17aG, F17bG, F17cG, UclD and UcaD. Structures were superimposed with $UcaD_{ntd}$ to generate the alignment. Residues are numbered according to their position in the native sequences. Identical residues are highlighted with red shading and white text, and similar residues with a blue frame and red text. The top and bottom secondary-structure depictions refer to $AtfE_{ntd}$ and $UcaD_{ntd}$ respectively. Residues that form side-chain interactions with bound GlcNAc in F17aG, F17bG and F17cG are marked with red stars.

Table 3
Top *DALI* hits for UcaD and AtfE.

(*a*) UcaD.

Protein	PDB code	R.m.s.d. (Å)	No. aligned	Identity (%)	*Z*
UclD	5nwp	1.0	193	84	32.1
F17c-G (GafD)	1oio	2.9	163	15	15.5
LpfD	5afo	3.1	144	11	9.8
MrkD1P	3u4k	3.7	136	13	9.0
CupB6	5cyl	3.1	146	11	8.5

(*b*) AtfE.

Protein	PDB code	R.m.s.d. (Å)	No. aligned	Identity (%)	*Z*
UclD	5nwp	2.2	163	19	17.5
F17c-G (GafD)	1oio	3.1	157	13	13.7
LpfD	5afo	3.1	138	12	9.9
MrkD1P	3u4k	3.4	134	13	8.9
CupB6	5cyl	3.3	142	15	7.5

C-terminal His tag of $UcaD_{211}$ takes on part of the role of the final (G) strand in the fold (Supplementary Fig. S1), suggesting that the domain border was trimmed too short and that the NTD in $UcaD_{211}$ is incomplete. Indeed, in the longer $UcaD_{217}$ construct the last six residues replace the His tag from the shorter construct to form the last β-strand and complete the Ig-like fold (Fig. 1). In the $UcaD_{211}$ structure, two cobalt-coordination sites were found in each asymmetric unit. One site is formed by His48 and two histidine residues from the His tag (His216 and His218). The second site is composed entirely of residues from the His tag: His213 and His215 in one protomer and His217 from the symmetry-related protomer (Supplementary Fig. S1). This explains why cobalt ions were essential for the crystallization of the short construct.

The $UcaD_{211}$ and $UcaD_{217}$ structures are very similar, with r.m.s.d.s of 0.594 Å for 176 equivalent C^{α} atoms in $UcaD_{211}$ and chain *A* of $UcaD_{217}$, and 0.763 Å for 179 equivalent C^{α} atoms in $UcaD_{211}$ and chain *B* of $UcaD_{217}$. The largest differences are found in the C-terminal region, as expected, as well as in the six first residues (21–26) and in the BC loop, which interacts with the N-terminal region. These latter differences were initially attributed to differences in crystal packing (the N-terminal region is involved in crystal packing between the two protomers in the asymmetric unit in the $UcaD_{217}$ structure), but are likely to be caused by an unfortunate accidental Tyr–His mutation at position 25 in the $UcaD_{217}$ construct that was discovered in the final stages of refinement and was confirmed by re-sequencing of the construct.

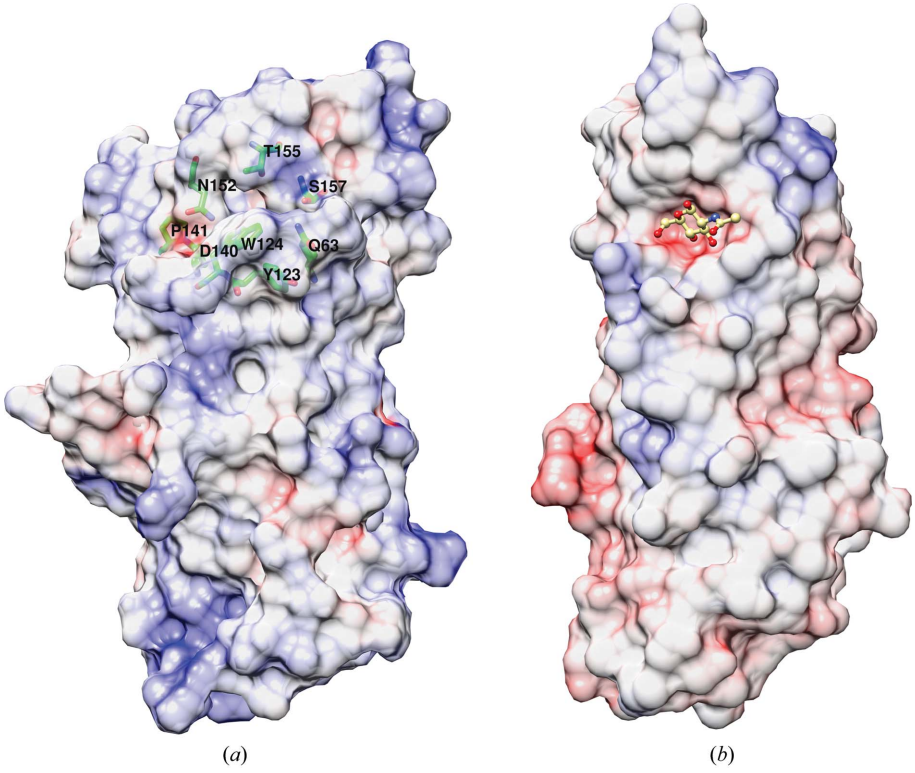


Figure 4
Comparison of the electrostatic surface potentials of $UcaD_{ntd}$ (*a*) and $F17cG_{ntd}$ (*b*). $UcaD_{ntd}$ and $F17cG_{ntd}$ are shown in surface representation with positively charged regions in blue, negatively charged regions in red and uncharged regions in white. The two structures are shown in the same orientation. (*a*) Residues in $UcaD_{ntd}$ forming a small pocket overlapping with the known GlcNAc-binding pocket in F17cG are highlighted as sticks. All residues are labelled and atoms are coloured according to element: carbon in green, oxygen in red and nitrogen in deep blue. (*b*) The known binding site of F17cG is indicated with the bound GlcNAc shown in ball-and-stick representation in pale yellow. O and N atoms are coloured in the same way as in (*a*).

$UcaD_{ntd}$ forms an elongated (approximate dimensions 60 × 35 × 30 Å) β-sandwich with the N- and C-termini at opposite ends of the structure (Fig. 1*a*). Based on the structures of other TDAs, the domain is expected to be oriented with its longest dimension along the direction of the fimbrial axis when incorporated into UCA fimbriae. Strands A1/2, G1/2, F1/2, C and D2 form one sheet of the sandwich, while strands B1/2, D1 and E form the second (referred to as the BED sheet below). The C–D and D–E loops (C′, C″; D′, α2) form a flap covering the BED sheet of the β-sandwich (Fig. 1).

A *DALI* search (Holm & Sander, 1995) for structural homologues of $UcaD_{ntd}$ identified the NTDs of F17G and F17G-like fimbrial adhesins (F17cG and UclD, respectively) as the top-scoring hits (Figs. 2*a* and 2*b*, Table 3). $UcaD_{ntd}$ shares 84% sequence identity with *E. coli* $UclD_{ntd}$ (Fig. 3) and hence might be expected to recognize a similar receptor. In contrast, the $UcaD_{ntd}$ sequence is only 15% identical to that of *E. coli* $F17cG_{ntd}$.

An O-linked oligosaccharide has been suggested as a possible UclD receptor based on the observation that

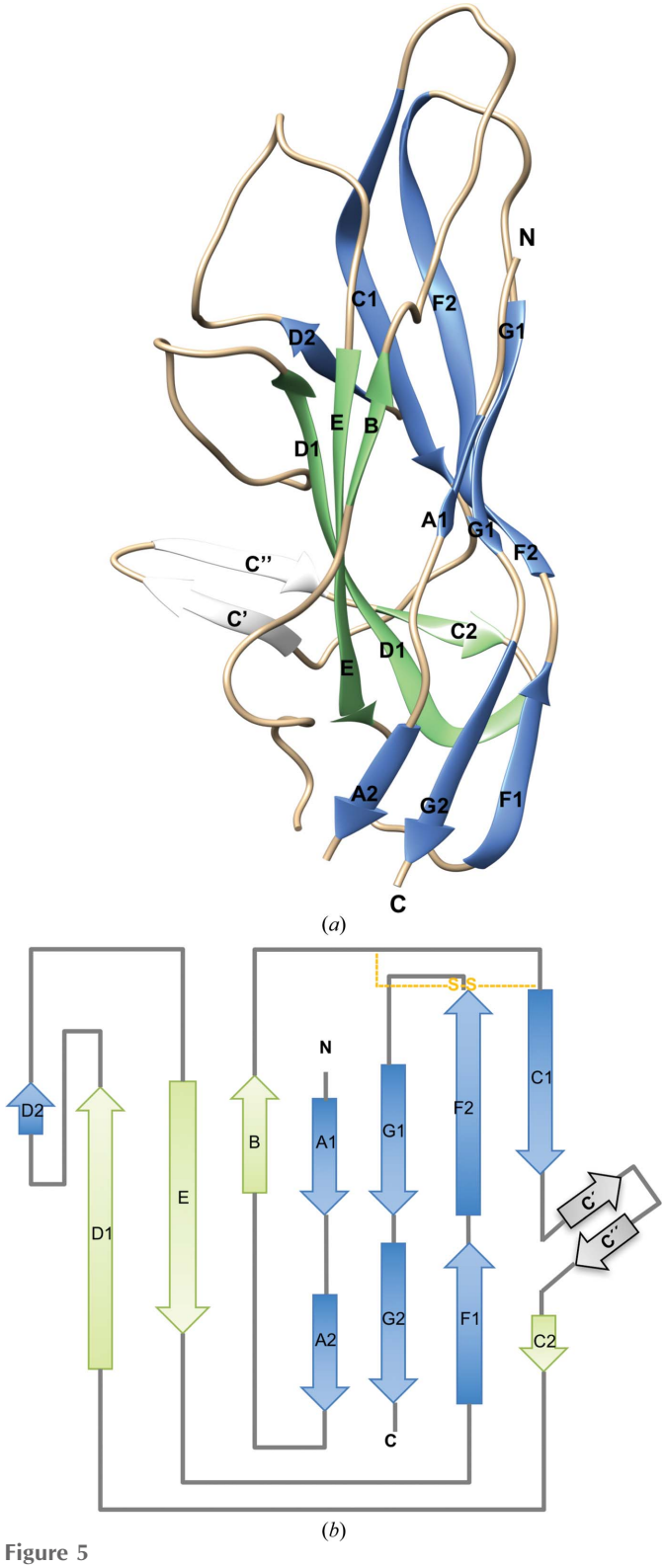


Figure 5
Three-dimensional crystal structure of $AtfE_{ntd}$. (*a*) Cartoon representation of $AtfE_{ntd}$. Secondary-structure elements are labelled A–G from the N-terminus to the C-terminus. β-Strands forming a β-sheet are in the same colour: β-strands A1/2, G1/2, F1/2, C1 and D2 are coloured blue, B, E, D1 and C2 are coloured green and C′ and C″ are coloured grey. Coils are coloured pale yellow. (*b*) Topology diagram of $AtfE_{ntd}$ with the secondary-structure elements coloured in the same way as in (*a*). A disulfide bond is highlighted in yellow with dashed lines indicating the positions of the two cysteines that form a disulfide bond in $AtfE_{ntd}$.

the binding of UclD in mouse colonic sections was abolished when pre-treating the tissue with O-glycosidase (Spaulding *et al.*, 2017). The region in UclD corresponding to the *N*-acetylglucosamine (GlcNAc)-binding pocket in F17cG (Buts *et al.*, 2003; Fig. 4) was also suggested as a potential receptor-binding site in UclD (Spaulding *et al.*, 2017), despite none of the residues being conserved (Fig. 3). The pocket found in this region in UcaD (and in UclD) is smaller and less charged than the F17cG pocket (Fig. 4). The pocket is lined with residues (Gln63, Tyr123, Trp124, Asp140, Asn152, Thr155 and Ser157) that could potentially hydrogen-bond to, as well as form π-stacking interactions with, a carbohydrate receptor. However, receptor-binding pockets are found in many different locations in fimbrial adhesins sharing the same overall Ig-like fold (De Greve *et al.*, 2007; Moonens *et al.*, 2012), although in most cases they are located in the top half of the NTD. Hence, in the absence of firm knowledge about the identities of the UclD and UcaD receptor(s), and given the lack of any significant sequence similarity between UclD/UcaD and F17cG, the prediction of a receptor-binding site based solely on a similar fold remains speculative.

Purified UCA fimbriae have been shown to bind to asialo-GM1, asialo-GM2 and lactosyl ceramide glycolipids *in vitro* (Lee *et al.*, 2000). However, since anti-UCA antibodies (which would be expected to chiefly target the major subunit present in fimbriae in huge excess over the single TDA per fimbria) blocked binding to the glycolipids, the observed binding is likely to be mediated by the major subunit rather than by UcaD. Also, while asialo-GM1 is commonly found on the surface of neuronal and immune cells, its presence on the surface of bladder epithelial cells has not been shown, casting doubt on the relevance of this finding for UTI.

3.2. Structure of $AtfE_{ntd}$

The $AtfE_{210}$ structure (Fig. 5) was solved by SAD phasing using a sodium iodide derivative. Despite the low sequence identity between $AtfE_{ntd}$ and $UcaD_{ntd}$ (18%; Fig. 3), the two structures are very similar (Fig. 2*c*), with an r.m.s.d. of 1.175 Å for 98 matched C^{α} positions. As for $UcaD_{ntd}$, the top *DALI* hits were UclD and F17cG (Table 3). Both $AtfE_{ntd}$ and $UcaD_{ntd}$ have distinct extended C′C″ insertions in the C–D loop compared with F17G TDAs (Figs. 1, 4 and 2*c*). Both structures have an AGFC sheet made from split stands, similar to F17G TDAs. $AtfE_{ntd}$ has a disulfide bond between Cys67 and Cys75 linking the end of strand B to the first residue in strand C. A conserved disulfide bond is present in F17G NTDs, but this connects the beginning of strand C to the DE loop. In $UcaD_{ntd}$ there is no disulfide bond. In contrast to F17G TDAs, both $AtfE_{ntd}$ and $UcaD_{ntd}$ have a *cis*-proline (Pro143 in AtfE and Pro141 in UcaD) at a structurally equivalent position right after strand D2. In UcaD (and UclD), this proline comes right after the asparagine residue (Asp140) in the pocket corresponding to the GlcNAc-binding pocket in F17cG (Fig. 4). Although *cis*-prolines often have specific structural and functional roles in proteins (Lorenzen *et*

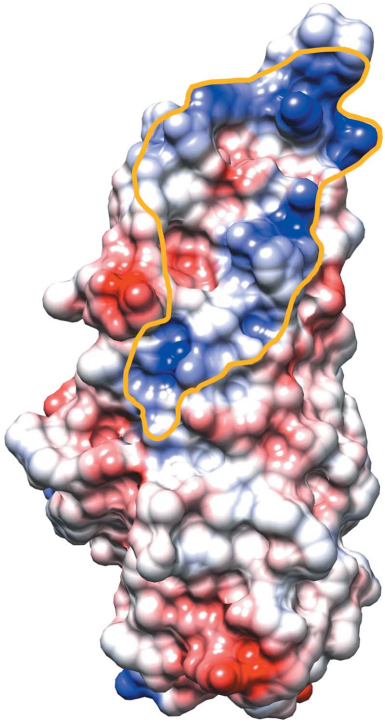


Figure 6
Electrostatic surface potential of $AtfE_{ntd}$. $AtfE_{ntd}$ is shown in the same orientation as $UcaD_{ntd}$ in Fig. 4. A positively charged ridge curving around the tip and extending approximately halfway down the domain, and surrounding two pockets, is outlined in orange. The upper pocket is located in roughly the same area as the GlcNAc-binding pocket in F17cG (Fig. 4).

al., 2005), the significance of the UcaD/AtfE *cis*-proline is unclear.

Studies have suggested ATF-dependent binding to exfoliated bladder sections rather than to intact bladder surfaces (Jansen *et al.*, 2004). Such regions typically expose the barrier layer of sulfated glycosaminoglycans (sGAGs) just below the superficial layer of umbrella cells lining the urothelium. The most important sGAGs in the bladder are heparan sulfate (HS), chondroitin sulfate (CS) and dermatan sulfate (DS), with CS located on the superficial (luminal) layer of the urothelium, while HS is found deeper down in the urothelium, in the basal membrane (Janssen *et al.*, 2013). Interestingly, during affinity chromatography of $AtfE_{210}$ we noted binding of the protein to the heparin matrix at both pH 5.5 and 8.0, suggesting not only nonspecific charge–charge interactions but also specific binding. Examination of the electrostatic surface of $AtfE_{ntd}$ reveals a distinct positively charged ridge curving around the tip and extending approximately halfway down the domain (Fig. 6). Two pockets are located within the borders of the ridge. It is tempting to speculate that this feature of AtfE might allow specific binding to sGAG structures such as, for example, HS, explaining the preferential binding to exfoliated bladder sections (Jansen *et al.*, 2004).

4. Conclusion

In summary, despite their divergent sequences, both $AtfE_{ntd}$ and $UcaD_{ntd}$ have similar structures. The structures, together with information from crystallization and protein-purification conditions, will provide the basis for future work on identifying receptors and the physiological functions of ATF and UCA.

Acknowledgements

We acknowledge the ESRF and DLS for the provision of synchrotron-radiation facilities and we would like to thank local contacts and beamline staff at ESRF beamlines ID23-1, ID23-2 and ID30A-3, and at DLS beamline I02, for assistance during data collection. We thank Jessica Schaffer for the preparation of genomic DNA, Sanjeewani Sooriyaarachchi for insights into protein purification, Terese Bergfors for constructive discussion in seeding experiments and Xiaohu Guo for help in collecting the $UcaD_{211}$ SAD data.

Funding information

We acknowledge funding from the Swedish Research Council (grant No. 2016-04451 to SDK) and the US Public Health Service (grant AI083743 to MMP).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Armbruster, C. E., Mobley, H. L. T. & Pearson, M. M. (2018). *EcoSal Plus*, **8**, https://doi.org/10.1128/ecosalplus.ESP-0009-2017.
Armbruster, C. E., Smith, S. N., Johnson, A. O., DeOrnellas, V., Eaton, K. A., Yep, A., Mody, L., Wu, W. & Mobley, H. L. T. (2017). *Infect. Immun.* **85**, e00808–16.
Aslanidis, C. & de Jong, P. J. (1990). *Nucleic Acids Res.* **18**, 6069–6074.
Bouckaert, J. *et al.* (2005). *Mol. Microbiol.* **55**, 441–455.
Buts, L., Bouckaert, J., De Genst, E., Loris, R., Oscarson, S., Lahmann, M., Messens, J., Brosens, E., Wyns, L. & De Greve, H. (2003). *Mol. Microbiol.* **49**, 705–715.
Cook, S. W., Mody, N., Valle, J. & Hull, R. (1995). *Infect. Immun.* **63**, 2082–2086.
De Greve, H., Wyns, L. & Bouckaert, J. (2007). *Curr. Opin. Struct. Biol.* **17**, 506–512.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
Geibel, S. & Waksman, G. (2014). *Biochim. Biophys. Acta*, **1843**, 1559–1567.
Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.
Jansen, A. M., Lockatell, V., Johnson, D. E. & Mobley, H. L. T. (2004). *Infect. Immun.* **72**, 7294–7305.
Janssen, D. A. W., van Wijk, X. M. R., Jansen, K. C. F. J., van Kuppevelt, T. H., Heesakkers, J. P. F. A. & Schalken, J. A. (2013). *J. Urol.* **189**, 336–342.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. (2015). *Nature Protoc.* **10**, 845–858.
Lee, K. K., Harrison, B. A., Latta, R. & Altman, E. (2000). *Can. J. Microbiol.* **46**, 961–966.
Lorenzen, S., Peters, B., Goede, A., Preissner, R. & Frömmel, C. (2005). *Proteins*, **58**, 589–595.
Massad, G., Bahrani, F. K. & Mobley, H. L. T. (1994). *Infect. Immun.* **62**, 1989–1994.

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.
Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C. & Ferrin, T. E. (2006). *BMC Bioinformatics*, **7**, 339.
Moonens, K., Bouckaert, J., Coddens, A., Tran, T., Panjikar, S., De Kerpel, M., Cox, E., Remaut, H. & De Greve, H. (2012). *Mol. Microbiol.* **86**, 82–95.
Nielubowicz, G. R. & Mobley, H. L. T. (2010). *Nature Rev. Urol.* **7**, 430–441.
Papadopoulos, J. S. & Agarwala, R. (2007). *Bioinformatics*, **23**, 1073–1079.
Pearson, M. M. *et al.* (2008). *J. Bacteriol.* **190**, 4027–4037.
Pellegrino, R., Scavone, P., Umpiérrez, A., Maskell, D. J. & Zunino, P. (2013). *Pathog. Dis.* **67**, 104–107.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.
Rocha, S. P. D., Pelayo, J. S. & Elias, W. P. (2007). *FEMS Immunol. Med. Microbiol.* **51**, 1–7.
Scavone, P., Iribarnegaray, V., Caetano, A. L., Schlapp, G., Härtel, S. & Zunino, P. (2016). *Pathog. Dis.* **74**, ftw033.
Schaffer, J. N. & Pearson, M. M. (2015). *Microbiol. Spectr.* **3**, 212–263.
Spaulding, C. N., Klein, R. D., Ruer, S., Kau, A. L., Schreiber, H. L., Cusumano, Z. T., Dodson, K. W., Pinkner, J. S., Fremont, D. H., Janetka, J. W., Remaut, H., Gordon, J. I. & Hultgren, S. J. (2017). *Nature (London)*, **546**, 528–532.
Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H. & Hung, L.-W. (2009). *Acta Cryst.* D**65**, 582–601.
Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 61–69.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wray, S. K., Hull, S. I., Cook, R. G., Barrish, J. & Hull, R. A. (1986). *Infect. Immun.* **54**, 43–49.
Zavialov, A. V., Berglund, J., Pudney, A. F., Fooks, L. J., Ibrahim, T. M., MacIntyre, S. & Knight, S. D. (2003). *Cell*, **113**, 587–596.
Zav'yalov, V., Zavialov, A., Zav'yalova, G. & Korpela, T. (2010). *FEMS Microbiol. Rev.* **34**, 317–378.
Zunino, P., Geymonat, L., Allen, A. G., Legnani-Fajardo, C. & Maskell, D. J. (2000). *FEMS Immunol. Med. Microbiol.* **29**, 137–143.