



Structural Insight into Host Recognition by Aggregative Adherence Fimbriae of Enteroaggregative *Escherichia coli*

Andrea A. Berry¹*, Yi Yang²*, Natalia Pakharukova³*, James A. Garnett², Wei-chao Lee², Ernesto Cota², Jan Marchant², Saumendra Roy^{3,4}, Minna Tuittila³, Bing Liu², Keith G. Inman⁵, Fernando Ruiz-Perez⁶, Inacio Mandomando⁶, James P. Nataro⁶, Anton V. Zavalov^{3*}, Steve Matthews^{2*}

1 Center for Vaccine Development, Department of Pediatrics, University of Maryland School of Medicine, Baltimore, Maryland, United States of America, **2** Centre for Structural Biology, Department of Life Sciences, Imperial College London, South Kensington, London, United Kingdom, **3** Department of Chemistry, University of Turku, Turku, JBL, Arcanum, Turku, Finland, **4** Department of Chemistry and Biotechnology, Swedish University of Agricultural Sciences, Uppsala BioCentre, Uppsala, Sweden, **5** Paragon Bioservices, Inc, Baltimore, Maryland, United States of America, **6** Department of Pediatrics, University of Virginia School of Medicine, Charlottesville, Virginia, United States of America

Abstract

Enteroaggregative *Escherichia coli* (EAEC) is a leading cause of acute and persistent diarrhea worldwide. A recently emerged Shiga-toxin-producing strain of EAEC resulted in significant mortality and morbidity due to progressive development of hemolytic-uremic syndrome. The attachment of EAEC to the human intestinal mucosa is mediated by aggregative adherence fimbria (AAF). Using X-ray crystallography and NMR structures, we present new atomic resolution insight into the structure of AAF variant I from the strain that caused the deadly outbreak in Germany in 2011, and AAF variant II from archetype strain 042, and propose a mechanism for AAF-mediated adhesion and biofilm formation. Our work shows that major subunits of AAF assemble into linear polymers by donor strand complementation where a single minor subunit is inserted at the tip of the polymer by accepting the donor strand from the terminal major subunit. Whereas the minor subunits of AAF have a distinct conserved structure, AAF major subunits display large structural differences, affecting the overall pilus architecture. These structures suggest a mechanism for AAF-mediated adhesion and biofilm formation. Binding experiments using wild type and mutant subunits (NMR and SPR) and bacteria (ELISA) revealed that despite the structural differences AAF recognize a common receptor, fibronectin, by employing clusters of basic residues at the junction between subunits in the pilus. We show that AAF-fibronectin attachment is based primarily on electrostatic interactions, a mechanism not reported previously for bacterial adhesion to biotic surfaces.

Citation: Berry AA, Yang Y, Pakharukova N, Garnett JA, Lee W-c, et al. (2014) Structural Insight into Host Recognition by Aggregative Adherence Fimbriae of Enteroaggregative *Escherichia coli*. PLoS Pathog 10(9): e1004404. doi:10.1371/journal.ppat.1004404

Editor: Jeremy P. Derrick, University of Manchester, United Kingdom

Received: March 9, 2014; **Accepted:** August 15, 2014; **Published:** September 18, 2014

Copyright: © 2014 Berry et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: AVZ is supported by the Finnish Academy (grant 273075; <http://sciencenordic.com/partner/academy-finland>). The EACEA (<http://eacea.ec.europa.eu>) supports NP for an Erasmus Mundus scholarship. SM is supported by the Wellcome Trust (Senior Investigator Award 100280, Programme grant 079819; equipment grant 085464; <http://www.wellcome.ac.uk>) and the Leverhulme Trust (RPG-2012-559; <http://www.leverhulme.ac.uk>). JPN and AAB are supported by a US Public Health Service grant (AI-033096; www.usphs.gov). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist. Although Keith G. Inman is employed by Paragon Bioservices Inc, this does not alter our adherence to all PLOS Pathogens policies on sharing data and materials.

* Email: anton.zavalov@utu.fi (AVZ); sj.matthews@imperial.ac.uk (SM)

† These authors contributed equally to this work.

Introduction

Enteroaggregative *Escherichia coli* (EAEC) was first identified in 1987 and it has since become recognized as a leading cause of acute and persistent diarrhea worldwide [1,2]. EAEC is characterized by its distinct “stacked-brick” pattern of aggregative adherence (AA) to HEp-2 cells *in vitro* [1]. This defining phenotype is mediated by aggregative adherence fimbriae (AAF). EAEC has been associated with persistent diarrhoea in children and in individuals infected with HIV [3,4]. It is also commonly detected in symptomatic travelers returning from developing countries [5].

Emergence of a Shiga toxin (Stx)-producing strain of EAEC [6,7] represents a significant threat to public health. The Stx-producing

O104:H4 strain responsible for the 2011 outbreak in Germany was notable for being more virulent than Shiga-toxin-producing *E. coli* strains that do not have virulence factors associated with EAEC; it resulted in 3816 cases of gastroenteritis, 845 cases of hemolytic uremic syndrome (HUS), and 54 deaths [8].

Most enteric bacterial pathogens possess specific adherence factors that are responsible for recognizing receptors on host cells prior to colonization. In the case of EAEC, there are four known AAF variants: AAF/I (encoded by the *agg* gene cluster), AAF/II (encoded by the *aaf* genes), AAF/III (*agg3*) and Hda/AAF/IV (encoded by the *hda* genes) [9–12]. All AAF fimbrial biogenesis genes are encoded on a 55 to 65 MDa plasmid called pAA. Different AAF variants are expressed by different EAEC strains, where they are required for the bacterium’s adherence to small

Author Summary

Enteroaggregative *Escherichia coli* (EAEC) is a major cause of diarrhea worldwide and is commonly present as an infection in symptomatic travelers returning from developing countries. The attachment of EAEC to the human intestine is mediated protein filaments extending from the bacterial surface known as aggregative adherence fimbria (AAF). Here we use X-ray crystallography and nuclear magnetic resonance (NMR) structures to provide an atomic structure of the protein fibers made by the two major variants, AAF/I and AAF/II. The structures of the major subunit proteins show that the AAFs assemble into flexible, linear polymers that are capped by a single minor protein subunit at the tip. Biochemical assays reveal that the AAFs recognize a common receptor, the extracellular matrix protein fibronectin, via clusters of positively-charged amino acid residues running along the length of the fimbriae. Our structures suggest a unique mechanism based on ionic interactions for AAF-mediated receptor binding and biofilm formation.

and large intestinal mucosa. AAF adhesins have also been shown to promote biofilm formation on abiotic surfaces (glass and plastic) [13]. Despite their shared phenotypes, AAF show significant differences in agglutination of erythrocytes from different species [12], suggesting that they may recognize different receptors or bind to the same receptors with different affinity.

AAF are assembled via the FGL chaperone/usher (CU) pathway [14,15] and share a similar gene cluster architecture with those of the Afa/Dr subfamily of polyadhesins [16,17]. As with Afa/Dr polyadhesins, AAF consist of two secreted protein subunits, a major subunit (A) and a putative minor subunit (B). The minor subunit of Afa/Dr polyadhesins may mediate invasion of host cells by uropathogenic *E. coli* [18]; usher-independent secretion has also been proposed [19,20]. In EAEC, the minor subunit of AAF/II (AafB) has been associated with the release of cytokines [10,21], suggesting its surface localization and function in EAEC pathogenesis. Whereas the minor subunits are relatively conserved within the subfamily, the major (A) subunits of AAF show only marginal sequence similarity among each other (Fig. 1) and no sequence similarity to the subunits from the Afa/Dr family (Fig. S1). Whereas most of the major subunits of CU-assembled fimbriae are typically negatively charged at physiological pH, the major subunits of AAF are positively charged up to pH 9.2–9.5; this feature is thought to play a role in EAEC adhesion. Moreover, a regulatory mechanism based on the repulsion of AAF and surface-localized dispersin has been proposed [22]. Although the major subunit of AAF/II has been recently shown to mediate attachment of EAEC strain 042 to the extracellular proteins fibronectin, laminin and type IV collagen [23], the mechanism of recognition is not known, and it is also not clear whether other types of AAF mediate attachment.

Previous transmission electron microscopy (EM) studies of the AAF expressed by *E. coli* showed the presence of thin, bundled fibers that can extend up to several microns from the bacterial surface. Estimates for the diameter of individual fimbriae have been reported to be in the 2–3 nm range for AAF/I [9,10], whereas images for AAF/II show fibers of up to 5 nm in width [10]. AAF/III are also long, flexible fimbriae with measured diameters of 3–5 nm, but these were usually observed as individual filaments and occasionally in bundles [24]. Based on homology of the CU machineries, it is predicted that these AAF structures are assembled by the FGL CU pathway in a similar fashion to the

Yersinia pestis F1 antigen [25], *E. coli* AFA-III [26], *E. coli* CS6 [27], *Yersinia pestis* pH6 antigen [28] and *Salmonella* Saf polyadhesins [29]. The major subunits from these systems lack the last (G) β -strand of a typical seven-stranded immunoglobulin fold, which exposes a substantial hydrophobic cleft running between the two β -sheets of the subunit. In the fiber, subunits are linked together by donor strand complementation (DSC) with an N-terminal G_d donor strand segment of one subunit inserted into the hydrophobic cleft of a neighboring subunit [25,30].

Theoretically, for any fimbrial subunit capable of DSC polymerization it is possible to create a circularly permuted construct [26,27,31,32]. In this construct, the G_d strand is placed at the C terminus of the subunit, enabling self-complementation and formation of a monomer with the classical Ig-fold. Whereas it is difficult to determine the structure of multimeric fimbriae at atomic resolution, production of stable monomers through donor strand complementation allows their study using nuclear magnetic resonance (NMR) spectroscopy and/or X-ray crystallography.

In this study, we report X-ray crystallographic and NMR studies of monomeric, donor-strand-complemented major and minor subunits of AAF/I from the German outbreak strain and AAF/II from EAEC strain 042 and provide an atomic resolution model for the structure of the entire fimbriae. Based on the structures and results of NMR titration experiments, site directed mutagenesis was performed to map the fibronectin binding sites in the organelles. Our results suggest that although there are significant structural differences between AAF/I and AAF/II, both uniquely rely on ionic-based mechanisms for adhesion to a common receptor.

Results

Model of AAF assembly and design of self-complemented subunits

Major subunits of AAF contain a pair of conserved cysteine residues that are commonly found in subunits of CU fimbriae (Fig. 1) [33,34]. As the first cysteine often marks the beginning of the subunit core structure, we hypothesized that the N-terminal sequences 12–24 residues preceding the cysteine in the major subunits could act as donor strands connecting subunits in the fiber. At the same time, based on the fact that minor subunits of AAF are 12–23 residues shorter than major subunits (Fig. 1), we hypothesized that they do not possess their own donor strand sequences and are stabilized in the fiber by strands donated by major subunits. A model, in which minor subunits insert at the tip of a polymer of major subunits, would be consistent with this hypothesis.

To verify this model, we engineered DSC-monomers for each of the major A and minor B subunits of AAF type I and II, encoded by the *agg* and *aaf* gene clusters in the German outbreak strain and archetypal strain 042, respectively, by extending them with the potential donor strand sequences of the corresponding major subunits (dsA): AggAdsA, AggBdsA, AafAdsA and AafBdsA (Fig. 2A and S2). To ensure that the donor strand has sufficient conformational freedom to insert correctly into the acceptor cleft, we introduced linker sequences between the subunit's last residue and the first residue in the donor strand. Using native signal sequences, AggAdsA and AggBdsA were expressed in the *E. coli* periplasm, whereas AafAdsA and AafBdsA were expressed in the *E. coli* cytoplasm and refolded. All four subunits were purified in soluble, monomeric form and exhibited high stability, suggesting that the chosen donor strands sequences efficiently stabilized the subunits.

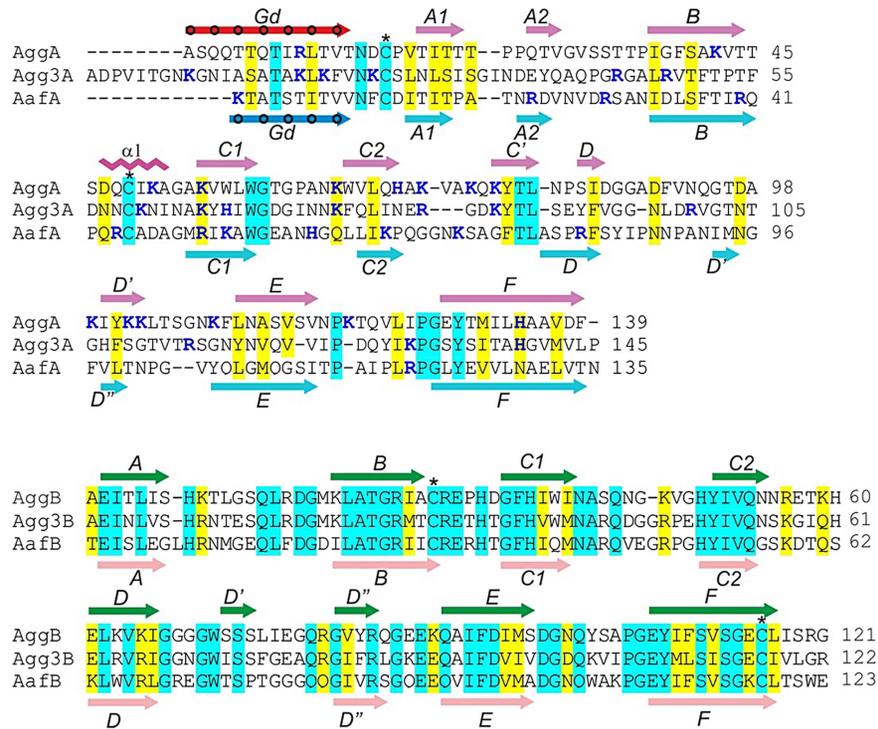


Figure 1. Sequence alignment of the major (AggA, AafA and Agg3A) and minor (AggB, AafB and Agg3B) subunits of aggregative adherence fimbriae (AAF) type I, II and III. Secondary structure elements of AggA, AafA, AggB, and AafB core structures are shown in magenta, cyan, green, and pink respectively, whilst the donor strands in AggA and AafA (G_d) are shown in red and blue, respectively. Donor residues occupying pockets of the acceptor cleft are indicated with circles. Amino acid identities and similar residues are indicated by background shading in cyan and yellow, respectively. The donor residues, occupying pockets of the acceptor cleft are indicated with circles. Positively charged residues are shown in bold and painted in blue. Cysteine residues involved in disulfide bonds are indicated with stars. CLUSTALW alignment of sequences was modified based on superposition of structures of the donor strand complemented (DSC) subunits AggA and AafA and AggB and AafB (this study). doi:10.1371/journal.ppat.1004404.g001

High-resolution structures of AggAdsA, AggBdsA, AafAdsA and AafBdsA

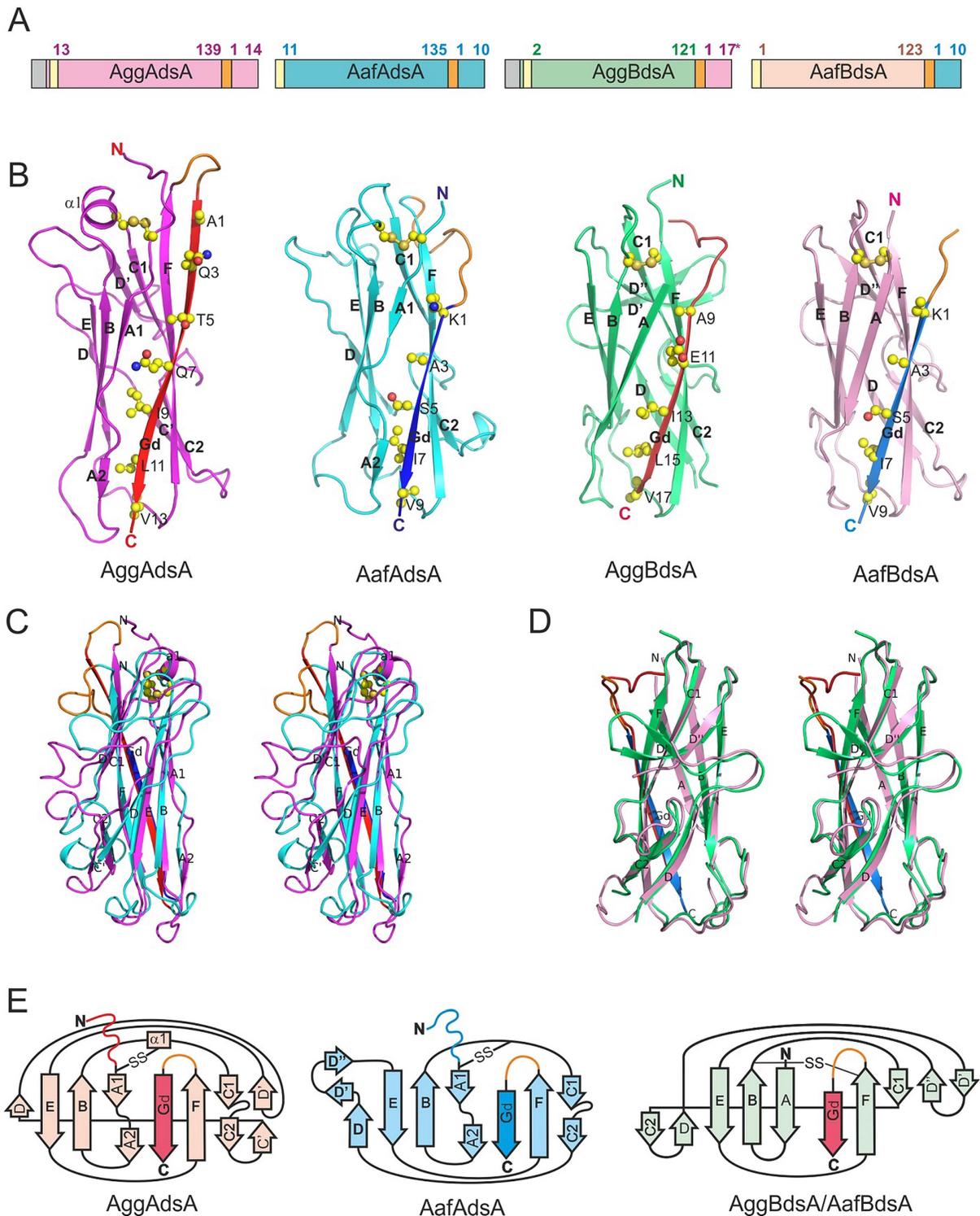
To elucidate the structure and assembly mechanism of AAF, we determined high-resolution crystal structures of AggAdsA (1.6 Å; pdb accession code 4PH8), AggBdsA (2.4 Å; pdb accession code 4PHX) and AafBdsA (3.0 Å; pdb accession code 4OR1) as well as the solution structure of AafAdsA (pdb accession code 2MPV and BMRB ID 25001) (Fig. 2 and Fig. S3 and Tables S1 and S2).

The DSC-monomers of the major subunits (AggAdsA and AafAdsA) have a classical Ig-like fold that consists of two β sheets packed against each other in a β sandwich (Fig. 2B). In solution, the sequence prior to the first conserved cysteine (residues 1–10, Fig S2) is highly flexible (Fig. S3A), suggesting that the N-terminal extension would be fully accessible for polymerization in preassembly monomers. On the other hand, when the N-terminal extensions are linked to the C-termini of the subunits to form the missing C-terminal G strand and facilitate self-complementation, they are well ordered and illustrate how major subunits form DSC contacts in AAF fibers.

Superposition of the AggAdsA and AafAdsA structures revealed variation at the edges of the β sandwich structure (Fig. 2C and E). In AafAdsA, a substantial strand D is hydrogen bonded to strand E of β sheet 1 (A₁BED) where as in AggAdsA, strand D is very short. Instead, it has additional strands C and D, which form the edge of β sheet 2 (CDC₁C₂FG_dA₂) in this subunit. Both strand F and the donor strand (G_d) at the other edge of sheet 2 are longer in AggAdsA than in AafAdsA. The donor strand of AggAdsA inserts seven classical donor residues in to the acceptor cleft (Fig. 2B),

which is two more than AafAdsA and other structurally characterized subunits from FGL CU systems. In addition, AggAdsA contains an α helix in the loop between B and C₁ strands, which is absent in AafAdsA. These features give AggAdsA a more elongated shape than AafAdsA.

The overall similarity between AggA and AafA, which are the closest structural homologues, displays an r.m.s.d. of 3.2 Å for 125 Cα atoms. The AFA/III major subunit AfaE (pdb: 2ixq) [35] was identified as the second most structurally similar protein to AggA (r.m.s.d. of 4.1 Å for 120 Cα atoms) and AafA (r.m.s.d. of 4.0 for 115 Cα atoms) by a DALI [36] search of the protein data bank, confirming the evolutionary relationship between AAF and AFA/Dr fimbriae. The maximal structural differences between the AAF subunits and AfaE were found in the same structurally variable segments: the loop region between strands C₂ and E, beginning of strand F and the ends of strands G_d and A₂, which include conserved Gly127 and Tyr129 residues in strand F and Leu11 in the donor strand G_d (residues are numbered according to the sequence of AggA) (Fig. S1 and S5). Another conserved feature represents the disulfide bond connecting the BC₁ loop (α helix in AggAdsA) and the beginning of the subunit fold. The donor strand in the AAF and AFA/Dr families shares the same interactions with its neighboring strand as in other FGL CU systems such as Saf [29] and the F1 antigen [25]. In these systems, the G_d strand lies at the sheet edge, whereas in the P-pilus and type 1 fimbriae it is shorter and sandwiched intimately between long A and F strands



[37]. The lack of a second pairing strand for G_d is a distinguishing feature of the FGL CU systems that is reflected in the requirement for the long G' strand in the chaperone [17].

The level of conservation of structural elements in major subunits generally correlates with the degree of their mobility in solution (Fig. S3). The conserved core structure of AafAdsA is considerably less mobile in the NMR ensemble than the structurally variable region including the sequence between strands D and E, the N-terminus, the BC₁ loop, and the beginning of strand G_d, all localized at the upper part of the molecule (Fig. S3B). To provide further characterization of mobility in these regions, heteronuclear NOE and T₂ relaxation data were recorded and analyzed. These data together with TALOS+ predictions [38] confirmed that the region between residues 89 to 96 is dynamic on the picosecond timescale. Not surprisingly, the engineered linker between the C-terminus and the G_d strand is also highly flexible.

Interestingly, many residues that are either conserved across both AAF/AFA families (Fig. S1) or conserved only within AAF major subunits (shaded blue in Fig. 1) are located in the structurally variable and flexible regions (Fig. S5). Specifically, four of these residues are highly surface exposed, namely the AAF specific basic residue position 55, Trp59, Thr80 and the AAF/AFA conserved asparagine/aspartate 97 (residues are numbered according to the sequence of AggA), (Fig. 3A insert, Fig. S5).

The minor subunits AggBdsA and AafBdsA have more similar structures than the major pilins: 117 matching C α atoms are superimposed with r.m.s.d. of 1.2 Å (Fig. 2B and D). The donor sequences derived from the corresponding major subunits form classical donor strands (G_d), each inserting five donor residues into the subunit core structure. The values of shape correlation statistic (S_c) [39] calculated for the acceptor cleft-donor strand interactions for AggBdsA and AafBdsA (0.733 and 0.756, respectively) were similar to those calculated for AggAdsA and AafAdsA (0.784 and 0.804, respectively). The good geometrical fit between the donor strands and acceptor clefts suggests that these interactions are native [27]. As predicted, the minor subunits have no N-terminal extension that could act as a donor strand: the first residue in the mature protein is structured and immediately followed by the first strand A. Hence, this structural evidence suggests that the minor subunits assemble by accepting donor sequences of the corresponding major subunits. As such sequences are only available in tip subunits of major subunit polymers, the AAF minor subunits are likely to be tip localized. This arrangement is consistent with EM localization studies of the AAF pilins. Bacteria lacking the major subunit are unable to assemble fimbrial surface structures and immunogold localization revealed the major subunit distributed along the fimbrial shaft [10]. Furthermore, the significant sequence identity with the AfaD tip invasin (>60%), which has been visualized as a tip subunit, together with the ability of AggB to functionally complement AfaD mutants are also suggestive of a similar localization for AAF minor subunits [35,40,41].

The topology of minor subunits differs from that of the major subunits in several other aspects (Fig. 2E). In minor subunits, strand A is not split between the β sheets, but locates entirely in the sheet 1. Instead, strand C₂ that in a classical Ig fold is hydrogen bonded to strand F, in minor subunits is hydrogen bonded to strand D and hence belongs to sheet 2. Strands D and D in sheet 2 of minor subunits have no analogues in the structure of major subunits. This region is the most dissimilar (Fig. 2C) and flexible (Fig. S3) in the structure. A conserved disulfide bond also stabilizes the minor subunits. However, this bond connects the BC₁ loop with the end of strand F and not the beginning of the fold as in major subunits. The distinct structure, high structural conservation

and specific localization suggest a highly specialized important function for the AAF minor subunits.

Atomic model of AAF

The structures of DSC monomers were used to model AAF (Fig. 3 and S6). The modules of the AAF shaft structures (-subunit:G_d-) were modeled based on AggAdsA or AafAdsA, while the corresponding terminal modules (subunit:G_d-) were modeled based on AggBdsA or AafBdsA. To revert the circular permutation of the DSC-monomers, the artificial linker sequences connecting C-termini of the DSC subunits with N-termini of donor strands were deleted and the C-termini of donor strands were bridged with N-termini of adjacent DSC subunits using native linker sequences TND and VNK in AggA and AafA, respectively. The fiber model was generated by assuming the same orientation between successive subunits as observed in the crystal structure of the mini-fiber of the F1 antigen [25].

To provide further evidence for these models we visualized AAF/II fimbriae purified from the prototypical enteroaggregative *E. coli* strain 042 using negative-stain transmission electron microscopy. Consistent with previous EM studies of bacterial-surface localized AAF fimbriae [10,11], our purified AAF fibers show a propensity to intertwine into bundles at high concentrations. However, upon dilution individual fimbriae can be clearly observed (Fig. 3C). These fibers are thin with diameter of 2.5–4 nm and highly flexible, occasionally exhibiting ~30° bends within relatively short stretches (20 nm, ~5 subunits). The extended appearance of these fibrillar structures and their inherent flexibility are consistent with the relatively small interdomain contact area present in our reconstructed models. Similar low resolution EM images have been obtained for the FGL-CU Saf fimbriae and subsequent reconstruction revealed thin, extended structures of ~2.5 nm width [42]. The dimensions of the observed fibers are also consistent with high-resolution structures for other FGL-CU organelles, namely the Caf1M-Caf1'-Caf1'' mini-fiber of the F1 antigen [25] and the AfaDE tip complex [35], which show very similar values for the intersubunit angle (Fig. S6).

The more elongated AggA produced a longer AAF/I fiber than that of AAF/II consisting of the same amount of shorter AafA subunits (Fig. S6). The edge of the fiber module that is involved in DSC (strands A and G_d) forms the less exposed side of the fiber helix. This region contains the majority of the conserved structural residues that are closely positioned and involved in assembly. The disulfide bond stabilizing the donor strand linker in major subunits functionally belongs to this region. The disulfide bond in minor subunits serves a different function in that it stabilizes the acceptor cleft (Fig. S7). The opposite segment of the structure (sequences between strands B and E) is more exposed. This mobile and structurally variable, yet conserved surface region is a potential candidate for a receptor binding site (Fig. 3A, insert).

Fibronectin is a common receptor for AAF/I and AAF/II

EAEC adheres to fibronectin, and in strain 042 the attachment is mediated by AAF/II [23]. To examine if fibronectin is a common receptor for AAF, we studied AggAdsA-fibronectin interactions using surface plasmon resonance (SPR) (Fig. S8). The experiment revealed specific binding with a dissociation constant of 16±2 μ M (Table 1). This value is similar to that previously found for AafAdsA [23]. Interestingly, AggBdsA also bound to fibronectin but with significantly lower affinity (~100 μ M). Whereas a single AggB is unlikely to promote attachment of bacteria to fibronectin, a polymer of AggA subunits would mediate a tight bacterial adhesion by establishing multipoint interactions with several fibronectin molecules.

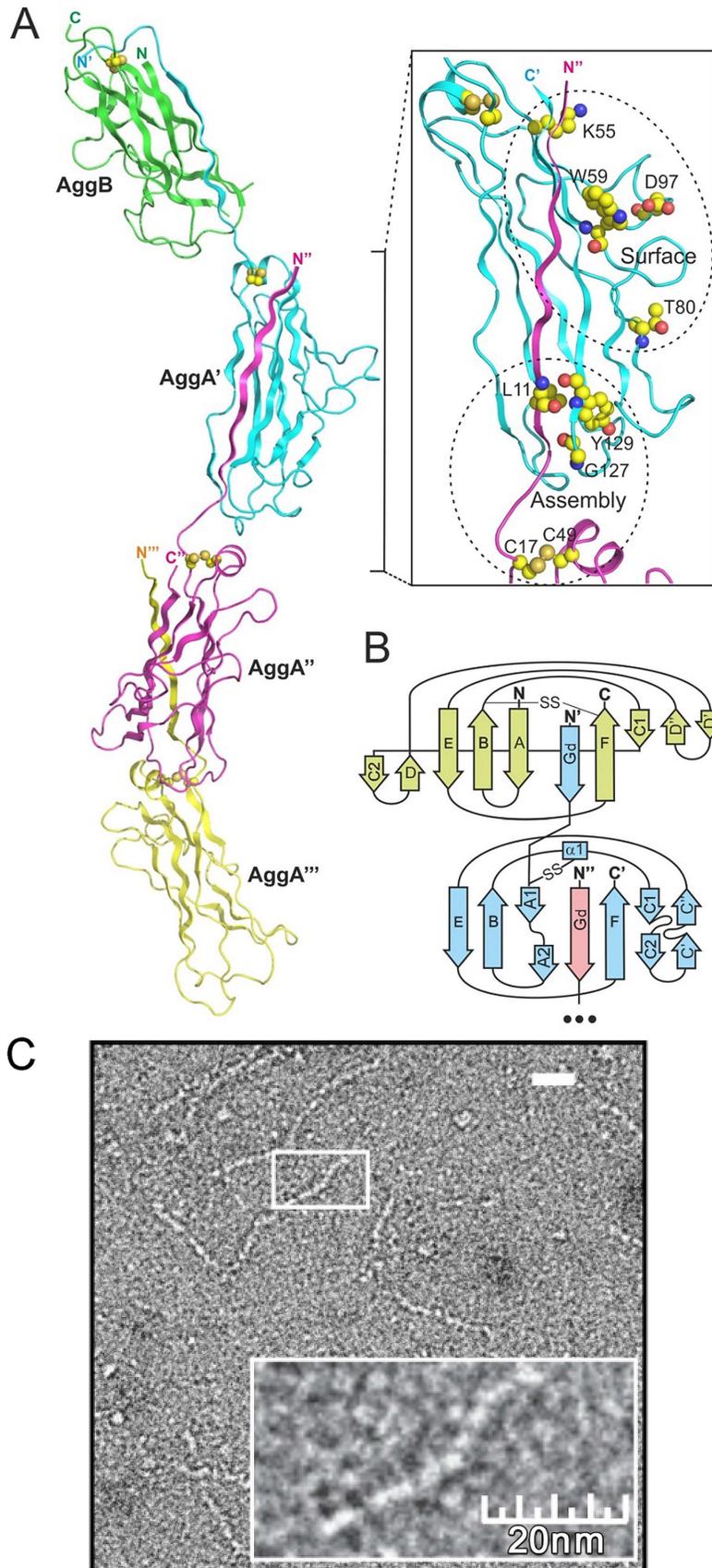


Figure 3. AAF architecture. (A) Model of AAF/I constructed based on the crystal structures of DSC subunits, AggAdsA and AggBdsA, and the crystal structure of the F1 antigen mini-fiber [25] (cartoon diagram). The fiber contains a single copy of the AggB subunit (green) at the tip of a polymer of the AggA subunits (a fragment containing three AggA subunits is shown). The insert shows localization of conserved residues in the structure of the fiber. (B) Topology diagram of the AAF/I fiber. (C) Negative stain transmission electron micrographs of diluted AAF/II fimbriae isolated from enteroaggregative *E. coli* strain 042.
doi:10.1371/journal.ppat.1004404.g003

AAF-fibronectin binding is mediated by ionic interactions

An NMR titration experiment was performed to identify the putative fibronectin-binding interface in AafAdsA. After comparing ^1H - ^{15}N HSQC NMR spectra of AafAdsA recorded in the absence and presence of FnI (Fig. 4A), several residues could be identified that exhibited significant chemical shift perturbations, suggesting the likely location of the binding surface (Fig. 4B). Unsurprisingly given the basic nature of AafA, a number of positive charged residues experienced chemical shift perturbations, namely Lys1, Arg40, Arg44, Arg51 and Lys53 and Lys72. Key residues in AafAdsA identified in the NMR FnI titration analyses were targeted for mutagenesis on the native fimbrial structure expressed in *E. coli*. Site-directed mutations were created in plasmid pBAD30 harboring the intact native *aafA* gene. Mutants were introduced into the 042 Δ *aaf* strain, and fibronectin binding in the complemented mutants was scored in the presence of arabinose (to induce *aafA* expression). A number of additional residues (Thr18, Arg23, Thr38, Lys66 and Thr114) were chosen, which are located in approximately equivalent regions to reported protein-binding interfaces for AfaE, namely carcinoembryonic antigen (CEA) and DAF [26,43,44]. Although AafA is not reported to bind to either CEA or DAF, it is conceivable that these surfaces on AafA are also important for fibronectin binding and/or biofilm formation. Protein expression was confirmed by Western blot (Fig. S9) and, perhaps with the exception of T18I, AAF production in the mutant strains is similar to wild-type. These expression data also imply that mutant subunits are stably polymerized into fimbriae. This is also reflected in surface exposure levels assessed qualitatively using immunofluorescence microscopy of AafA mutants Ser30Ala, Arg40Ala, Arg44Ala, Lys66Ala, Lys72Ala, and Thr77Ala (Fig. 5C). It, however, cannot be ruled out that some variation in AAF presentation may contribute to the fibronectin binding or biofilm phenotypes.

AafA mutant Arg40Ala exhibited significant reduction of biofilm formation (Fig. 5A), while mutants Arg23Ala, Arg40Ala,

Arg44Ala, Lys66Ala, and Lys72Ala displayed reductions in fibronectin binding compared with wild type (Fig. 5B). All other mutations did not affect either biofilm formation or Fn binding, including conservative substitutions Lys66Arg and Lys72Arg, which did not affect Fn binding. These data confirmed the identity of several residues suggested by NMR spectroscopy, and implicated two regions at the poles of the AafA subunit. Our data also suggested that the predicted receptor-binding site for DAF in the Dr adhesins is not involved in Fn binding. Interestingly, all residues identified as functional in fibronectin binding were basic arginines or lysines. As the high content of basic residues is a prominent feature for AAF, we hypothesized that basic residues are important for the AAF/I-fibronectin interactions. As EM studies of both bacterially-displayed and sheered AAF revealed an abundance of bundled fibers, it is also possible that there is a significant electrostatic contribution to this morphology and these intertwined bundles are important for function. This is consistent with the observation of a strong dependence of the strength of AggAdsA-fibronectin binding on the ionic strength of the solution (Table 1, Fig. S8). The affinity dropped by more than 12 fold in presence of 300 mM NaCl in 50 mM Hepes buffer, which could be explained by disruption of salt bridges between the interacting molecules.

To verify this hypothesis, we reduced the charge on the AggA surfaces that correspond to the equivalent basic regions identified in AafA by mutating groups of positively charge amino acids to alanine and measuring the interaction of the self-complemented monomer (i.e. AggAdsA) with fibronectin using SPR (Table 1). We focused the mutagenesis on surface patches comprising proximal pairs of lysines (including Lys51 and Lys109, Lys55 and Lys103, and Lys73 and Lys76) and the three closely positioned lysines Lys73, Lys76, and Lys78. All the double mutants exhibited significant decreases in fibronectin binding and, despite the absence of the cooperatively expected for a polymeric AggA fiber, the affinity for the triple mutant of monomeric AggAdsA dropped below the detection limit of the experiment. To confirm the structural integrity and stability of these charge removal mutations, the AggAdsA triple mutant (Lys73Ala, Lys76Ala, and Lys78Ala) was characterized further by circular dichroism (CD). The CD spectra of both WT and mutant monomeric proteins are almost identical showing a negative band at 217 nm and positive band at 195 nm (Fig. S10A). The CD profiles are characteristic of a highly similar β -sheet structure and thermal denaturation revealed that the mutations do not affect stability of the monomeric AggAdsA subunit (Fig. S10B). The defect in fibronectin binding for the mutant is therefore due exclusively to the absence of specific positively charged side chains. The low affinity and electrostatic nature of this interaction raises the question that AAF may target acidic proteins nonspecifically. We examined this using SPR and show that this is unlikely, as binding of AggAdsA to fibronectin is unaffected by the presence of 6-fold molar excess of Bovine serum albumin (BSA pI = 4.7; Fig. S11).

The conserved surface tryptophan residue in AAF major subunits (Trp59 and Trp55 in AggA and AafA, respectively) is located close to the conserved positive charges implicated in the AggA-fibronectin binding (Fig. 3A). In addition, AggA contains a closely positioned surface Trp57 as well as the exposed Phe91 and

Table 1. Effects of ionic strength and amino acid substitutions on the dissociation constant for AggAdsA-fibronectin binding.

Mutation	K_D , μM
WT, 75 mM NaCl	16 \pm 2
WT, 300 mM NaCl	>200
Trp57	31 \pm 6
Trp59	17 \pm 3
Phe91	20 \pm 3
Arg10	23 \pm 3
Lys51 and Lys109	49 \pm 4
Lys55 and Lys103	56 \pm 5
Lys73 and Lys76	61 \pm 6
Lys73, Lys76, and Lys78	>200

doi:10.1371/journal.ppat.1004404.t001

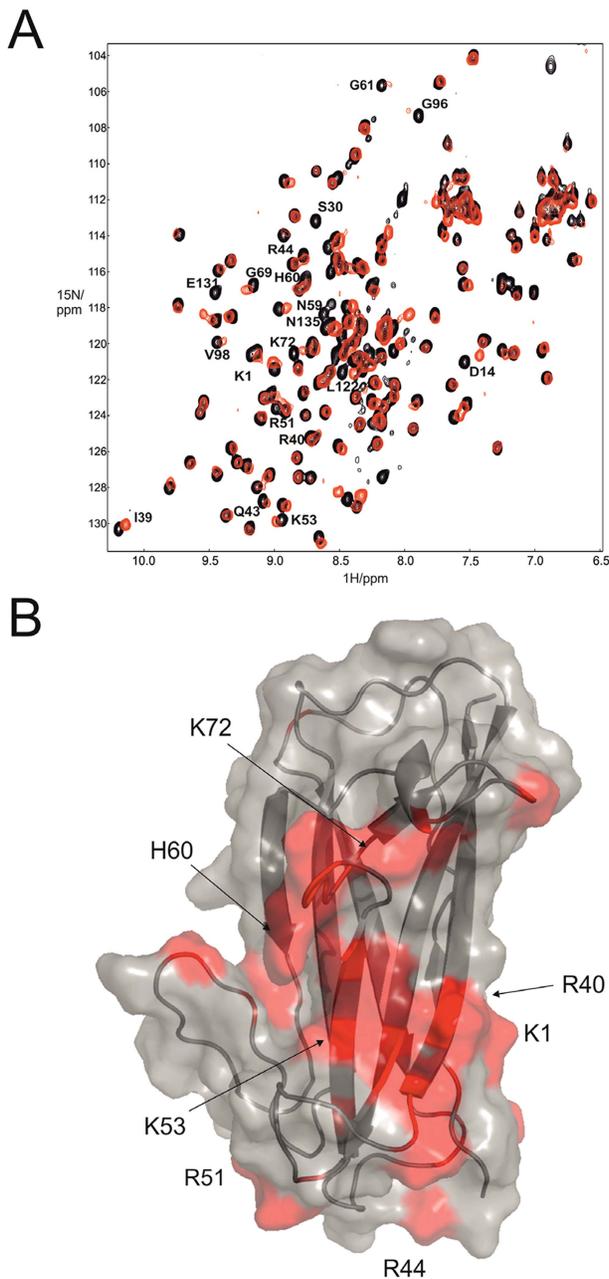


Figure 4. Fibronectin binding site mapping using NMR. (A) Overlaid ^1H - ^{15}N HSQC NMR spectra for free AafAdsA (black) with two molar equivalents of fibronectin (red). Residue labels represent mature protein sequence numbering. (B) Map of chemical shift perturbations on the structure of subunit. Key basic residues are highlighted. doi:10.1371/journal.ppat.1004404.g004

partially exposed Ile85, located in the variable region. To study the possible involvement of the extensive hydrophobic surface of these residues in AggA-fibronectin binding, we mutated each residue to produce the Trp57Ala, Trp59Ala, Ile85Ala and Phe91Ala mutants of the monomeric AggAdsA subunit. Trp57Ala, Trp59Ala, and Phe91Ala, but not Ile85Ala, were expressed, demonstrating that they are not essential for the protein structure. Examination of the fibronectin-binding of these AggAdsA mutants with the SPR binding assay revealed affinities similar to that of the wild type subunit (Table 1), suggesting that the large hydrophobic

surfaces displayed on AAF subunits are not involved in the fibronectin binding. Furthermore, the prominence of basic lysine and arginine residues suggests that fibronectin binding by AAF/I and AAF/II is driven by electrostatic interactions.

Modeling of fibronectin binding sites on AAF

Superimposition of fibronectin-binding residues on the fiber models placed the experimentally determined fibronectin-binding pocket within the clefts formed by adjacent subunits in the fiber (Fig. 6). The pocket is particularly well pronounced in AAF/I. Three closely positioned lysines (73, 76 and 78), mutations of which practically abolished the binding, are located in the loop between the C_2 and C strands at the bottom of the subunit. The other binding residues are located close to the top of the molecule: lysines 51 and 55 in the helix within the BC_1 loop and at the beginning of strand C_1 , respectively, and lysines 103 and 109 at the end of strand C and in the CE loop, respectively. These two polar segments are closely positioned in the fiber and form a continuous surface, which is characterized with the highest positive potential and contains a shallow groove in the middle. Similarly, in AAF/II all implicated binding residues locate at the junction between subunits with the fiber. Three arginines; 23 (end of the strand A_2) at the bottom of AafA; 40 (beginning of strand B); and 44 (the BC_1 loop) at the top of AafA are a part of the AAF/II surface with the highest positive potential. Interestingly, arginines 23, 40, and 44 form an array along the fiber length. Lysines 66 and 72 form an independent cluster locating at the subunit interface. It is likely that these residues contribute to fibronectin binding by forming additional contacts, which likely result from changes in the range of relative domain orientations along the fiber due to its flexibility. Alternatively, they might contribute to binding indirectly, either by promoting fiber bundling or stabilizing a binding competent conformation of the whole fiber. Although the majority of the identified binding residues are not conserved at specific sequence position (Fig. 1), many of them locate in the structurally similar regions at the poles of the subunits, which in the context of AAF fibers positions a high concentration of basic residues in the cleft between subunits.

Discussion

The AAF of EAEC is a multifunctional organelle that contributes multiple phenotypes plausibly related to pathogenesis. AAF have been associated with adherence to human intestinal explants [45], biofilm formation on abiotic surfaces [13] and adherence to polarized T84 cell monolayers with stimulation of IL-8 release and opening of epithelial tight junctions [46]. Although no single receptor has yet been associated directly with these phenotypes, binding to extracellular matrix proteins has been associated with colonization and biofilm formation. The AAF adhesins can be classified into at least four variants [12] with conserved characteristics, despite low levels of amino acid identity among the major pilin subunits. In this work, we present the structures of the major and minor pilin subunits of the most common AAF variants, AAF/I and AAF/II. We find that despite less than 25% amino acid identity between the AAF/I and AAF/II major pilins, AggA and AafA respectively, the two pilin proteins adopt a similar fold with shared surface characteristics. The most notable feature of the AggA and AafA structures is the unusual surface distribution of basic residues, giving rise to subunits with a substantial net positive charge at physiologic pH. Although the majority of bacterial adhesins are not highly positively charged, the adhesin of *Stenotrophomonas maltophilia* is basic and has also been associated with biofilm formation and adherence to

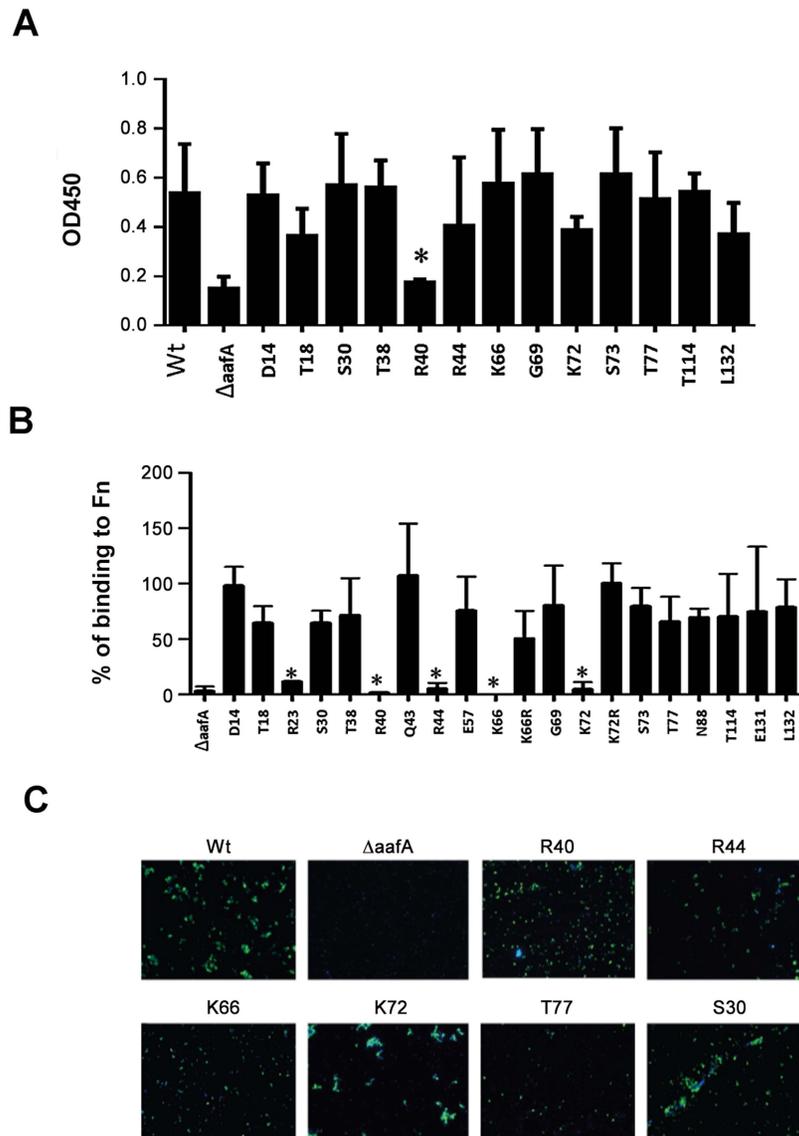


Figure 5. Involvement of AafA and site-directed-mutants in biofilm formation and binding to fibronectin. (A) Biofilm quantification. Bacteria were cultivated in LB for 20 h at 37°C in 24-well dishes and induced with 0.2% arabinose. Biofilms were fixed and stained with crystal violet, and then the stains were solubilized and quantitated spectrophotometrically at 450 nm. The bars represent the means of the results from triplicate wells; error bars indicate one standard deviation. Wt, wild type EAEC042, $\Delta aafA$, EAEC042 $aafA$ mutant. All residues tested were mutated to Ala except Thr18Ile and Thr38Leu. *, $P < 0.005$. (B) Fibronectin binding. EAEC 042 derivatives harboring site mutations in *aafA* were added to 96-well plates coated separately with 25 $\mu\text{g/ml}$ of fibronectin, and the binding was determined by collecting the cells adhered to wells by scraping them into PBS. The number of adherent bacteria was determined by counting the resulting colonies in duplicate and normalized to the wild type adherence (as a 100% adherence). The bars represent the means for three experiments, with the error bars indicating 1 standard deviation. *, significantly different from EAEC042WT ($P < 0.005$). (C) AafA surface expression. After induction of bacterial strains with 0.2% arabinose, cells were harvested, washed twice with PBS, and incubated with a polyclonal anti-aafA antibody (2 $\mu\text{g/ml}$) in PBS plus 1% bovine serum albumin for 1 h at room temperature with agitation. They were washed twice with PBS and incubated with a goat anti-rabbit IgG-fluorescein isothiocyanate (FITC) conjugate (10 $\mu\text{g/ml}$) for 30 min at room temperature in the dark. The washings were repeated, and the samples were resuspended in DAPI (4 mg/liter) and spread on slides for viewing by fluorescence microscopy. doi:10.1371/journal.ppat.1004404.g005

indwelling vascular catheters [47]. The contribution of positively charged adhesins to pathogenesis for either *Stenotrophomonas* or EAEC has largely been overlooked.

We present a new structural insight into the architecture of AAF/I and AAF/II fimbriae and reveal a contiguous basic surface located at the subunit interfaces within the biopolymer. Although the observation of highly basic AAFs in bundles is somewhat surprising due to potential repulsion between entwined fimbriae,

intersubunit repulsion between subunits within a fiber may enable these long, thin fimbriae to project from the bacteria surface as far as possible and prevent them from collapsing onto it. The alignment of AAFs into bundles may be a consequence their extended length and inherent helicity. There is precedent in other biological fibers from bacteria; for example, in the Type IV pilus from *N. gonorrhoeae* extensive phosphorylation does not abrogate bundle formation, but reduces the curliness of fibers and thickness

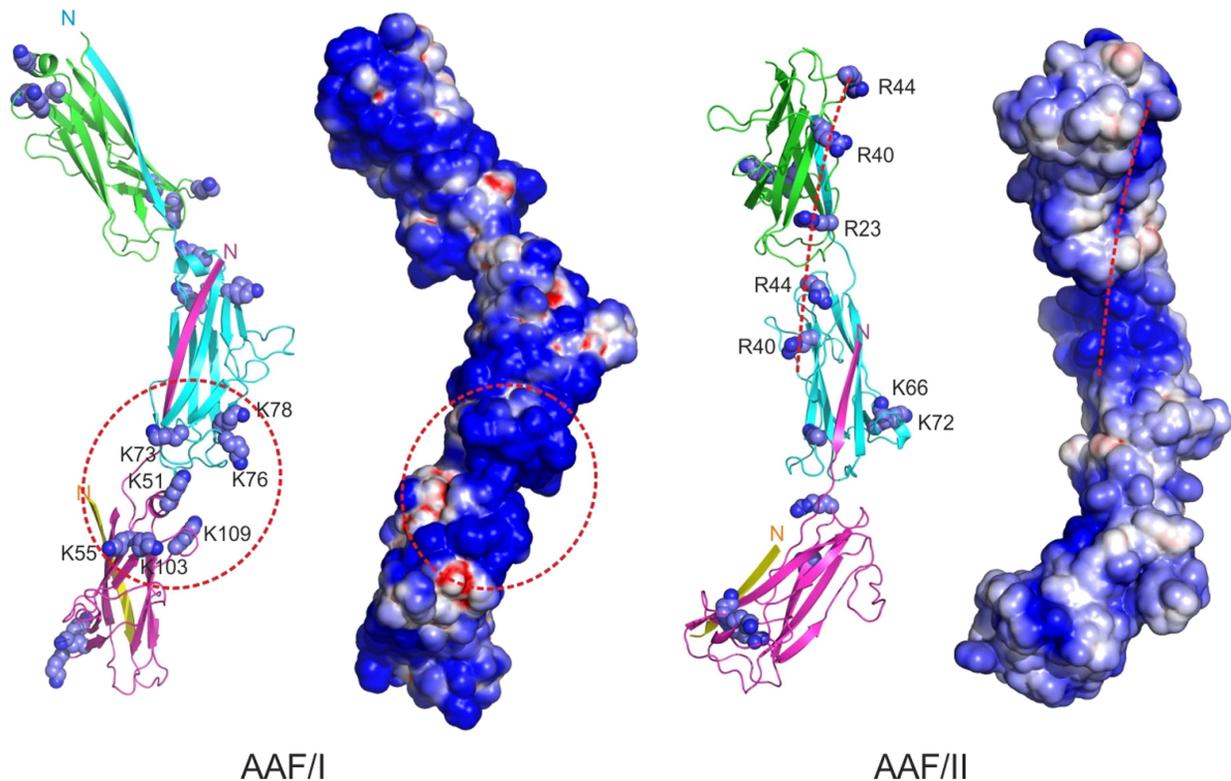


Figure 6. Location of fibronectin binding sites in AAF type I and II. Cartoon diagrams (left) and electrostatic potential assessable surfaces (right) of three major subunit polymers are shown. Positively charged residues interacting with fibronectin are drawn as sphere models. Arg40 in AAF/II is also involved in biofilm formation. Regions of positive surface potential are depicted in blue and negative potential in red (scale is from -4 to 4 kT/e). AAF/II contains two fibronectin binding surfaces: a continuous array of arginines 23, 40, and 44 and a repeated patch of lysines 66 and 72. In AAF/I, the basic residues concentrated at the interface between adjacent subunits. Seven lysine residues 51, 55, 73, 76, 78, 103 and 109 in this region are involved in fibronectin binding.
doi:10.1371/journal.ppat.1004404.g006

of bundles [48]. Furthermore, counterion coating of highly-charged surface of nanometer fibers has been shown to promote bundle formation [49].

We show that the basic nature in AggA and AafA plays an important role in the specific recognition of fibronectin. The mode of recognition is principally by electrostatic interaction, which is in contrast to mechanisms utilised by other bacterial fibronectin-binding proteins. Perhaps the best characterised structurally are the MSCRAMMs (microbial surface components recognizing adhesive matrix molecules) from Gram positive bacteria. MSCRAMMs are large modular proteins covalently attached to the bacterial cell wall. Examples of this family from staphylococci and streptococci bind fibronectin with high affinity through interaction with an extensive region with several copies of a repeat sequence. Structural studies have revealed that the repeats are disordered in isolation, but in complex they augment the β -sheets of fibronectin type I domains, forming an extensive β -zipper along the edge of consecutive domains, resulting in an extremely tight interaction with nM dissociation constants [50]. While the dissociation constants for single pili lie in the μ M range, the high avidity from the polymeric architecture of AAF would equate to a much stronger interaction with fibronectin. Although the precise role of fibronectin binding in EAEC pathogenesis is unclear, the strength of interaction and conserved characteristics between AggA and AafA fibronectin binding suggest a significant involvement. This also consistent with the persistent nature of EAEC infection which implies that bacteria progress beyond

simple apical attachment on the epithelial surface to maintain disease. Indeed, while extracellular matrix proteins are usually localized to the remote basement membrane, access to a bacterial pathogen can occur during inflammation, invasion or specific breaching of tight junctions. Tight junction opening has been associated with AAF expression during the EAEC infection of T84 cells [46]. This may be the consequence of a colonization signal cascade in which proinflammatory cytokines are secreted from infected epithelial cells [21]. The receptor associated with this phenotype has yet to be identified, however we expect that our structural insight into the electrostatically driven mechanism for AAF/II binding provides new inspiration for mutagenesis experiments that may accelerate this effort. It is also plausible that a natural receptor for AAF adhesins is not fibronectin itself, but another or multiple extracellular protein(s). In addition to fibronectin, other host receptors have been implicated in EAEC adhesion [51] and more recently both laminin and cytokeratin8 (CK8) have been confirmed to interact with the major subunit of AAF/II fimbriae [52]. It is conceivable that AAF have evolved a more generalised electrostatic mechanism for binding several host receptors that have appropriate arrangements of acidic residues, negatively-charged glycosylation sugars and/or phosphorylated site.

We have also determined the structures of the minor pilin subunits AggB and AafB. As expected from the similar major pilin subunit structures and a higher degree of sequence conservation between the minor pilin subunits, the structures of the two minor

pilin proteins are much more conserved. Although the role of the minor pilins in AAF is unknown, it has been suggested [17], and our data support this model, that the minor subunit caps the end of the AAF structure. Bacterial fimbriae are in general highly immunogenic, and antibodies raised against the structures are characteristically protective against bacterial colonization [53–55]. As such, fimbriae are high-profile vaccine candidates, and structural information can advance vaccine development efforts dramatically [56].

Materials and Methods

Construction of donor strand complemented monomers

The AggAdsA and AafAdsA expression constructs (pET101D-AggAdsA-O104H4 and pQE30-AafAdsA, respectively) were created as described previously in [57] and [58], respectively. The sequence coding for the *aggBdsA* gene was ordered from GenScript and placed under T7 promoter of pET101D to create the expression plasmid pET101D-AggBdsA. The sequence coding for the *aafBdsA* gene was ordered from Invitrogen and placed under T7 promoter of pQE20 to create the expression plasmid pQE30-AafBdsA.

Protein preparation

The *aafAdsA* and *aafBdsA* constructs were transformed into *E. coli* strain M15 containing the pREP4 plasmid (Qiagen). Cells were grown at 37 °C in either LB or M9 minimal medium supplemented with ¹⁵NH₄Cl and/or ¹³C-glucose (Cambridge Isotope Laboratories) and expression induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD_{600nm} = 0.6. Cells were harvested after 4 hrs, resuspended in 50 mM sodium phosphate pH 8.0, 8M urea, 300 mM NaCl, lysed using a French press, before being purified under denaturing conditions with Ni-NTA. The eluates were first dialysed against 50 mM sodium acetate pH 5, 50 mM NaCl, 1 M urea, followed by a second dialysis against the same buffer with no urea and finally gel filtered using a Superdex 75 gel-filtration column (GE Healthcare).

AggAdsA was expressed in the periplasmic space of *E. coli* strain BL21-AI, extracted by osmotic shock and purified as described in [57]. AggBdsA was expressed and extracted using the same procedure [57]. To remove the majority of contaminating proteins, the extract was filtrated through a 20-ml Source 30Q column (GE healthcare) in 20 mM Tris-HCl, pH 8.5. The sample was dialyzed overnight in 20 mM HEPES, pH 7.2 buffer, and purified further by cation-exchange chromatography on a Mono-S 5/50 GL column (GE healthcare) using a 0–250 mM elution gradient of NaCl. To obtain highest purity samples, protein was subjected to gel-filtration on a Superdex 75 column (GE Healthcare) equilibrated with 50 mM HEPES, pH 7.5 and 150 mM NaCl. Protein was concentrated to 33 mg ml⁻¹ for crystallization experiments on a Vivaspin device (GE healthcare) with molecular weight cut-off of 5 kDa.

Crystal structure determination

Crystallization of AafBdsA was performed by “sitting-drop” vapor-diffusion method grown in 0.1 M Bis-Tris pH 5.5, lithium sulphate and 25% w/v PEG 3350. Crystals were soaked for 30–60 s in cryoprotection solution (well solution complimented with 20% PEG 400) and then cooled by plunging them into liquid nitrogen. Diffraction data were collected under liquid-nitrogen cryoconditions at 100K on beamline I24 at the Diamond Light Source (DLS), UK. Data were processed with XDS [59]. The protein structure was solved by molecular replacement method using DraD protein (pdb accession code 2AXW: [60]) as a model

using Phaser [61] and refined with Refmac [62]. Coordinates have been deposited with the protein databank with accession code 4OR1.

Crystallisation and the quality of preliminary diffraction data for AggAdsA have been described previously [57]. Crystallization of AggBdsA was performed by “sitting-drop” vapor-diffusion method using commercial screening kits Index-HR2-144, JCSG+ Suite (Qiagen) at 290 K. Crystals grew in drops with 0.2 M Lithium sulfate monohydrate, 0.1 M Tris HCl pH 8.5, 30% w/v PEG 4000. Crystals were soaked for 30–60 s in cryoprotection solution prepared by mixing two parts of precipitant solution with one part 50% PEG 400 and then cooled by plunging them into liquid nitrogen. Diffraction data were collected under liquid-nitrogen cryoconditions at 100K on beamline ID23-1 at the European Synchrotron Radiation facility (ESRF), Grenoble, France. Data were processed with XDS [59]. The protein structure was solved by molecular replacement method using DraD protein as a model (55% of sequence identity) using Phaser from the PHENIX Software package [61].

NMR structure determination

For AafA-dsA, backbone and side-chain assignments were completed using our in-house, semi-automated assignment algorithms and standard triple-resonance assignment methodology [63]. H_α and H_β assignments were obtained using HBHA (CBCACO)NH. The side-chain assignments were completed using HCCH-total correlation (TOCSY) spectroscopy and (H)CC(CO)NH TOCSY. Three-dimensional ¹H-¹⁵N/¹³C NOESY-HSQC (mixing time 100 ms at 800 MHz) experiments provided the distance restraints used in the final structure calculation. The ARIA protocol [64] was used for completion of the NOE assignment and structure calculation. The frequency window tolerance for assigning NOEs was ±0.04 ppm and ±0.06 ppm for direct and indirect proton dimensions and ±0.6 ppm for both nitrogen and carbon dimensions. The ARIA parameters p, Tv and Nv were set to default values. 144 dihedral angle restraints derived from TALOS were also implemented [65]. The 10 lowest energy structures had no NOE violations greater than 0.5 Å and dihedral angle violations greater than 5°. Although structure calculations readily converged without the introduction of manual assignments, a systematic check of automatically-assigned NOEs was carried out. The structural statistics are presented in Table S1.

NMR titrations

¹H-¹⁵N TROSY/HSQC spectra of ¹⁵N-AafA-dsc was recorded at 25μM in the absence and presence of the 30-kDa N-terminal type I repeat domain of fibronectin (FnI) in equimolar concentrations.

Construction of AafA and AggA mutants

Bacterial strains were grown in Luria-Bertani (LB) broth at 37°C with shaking unless otherwise indicated. When appropriate, the medium was supplemented with antibiotics at the following concentrations: ampicillin, 100 μg/ml and kanamycin 50 μg/ml. The bacterial strains used in this study were DH5α [supE44 ΔlacU169 (φ80 lacZAM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1], EAEC042WT, and EAEC042 [10].

Construction of pBADaafA plasmid

Inducible AafA expression was achieved by cloning *aafA* into the pBAD30 plasmid. To stabilize AafA expression, its chaperone *aafD* was also cloned bicistronically into the same plasmid. Briefly,

a DNA fragment containing *aafD* and *aafA* was amplified by PCR from EAEC 042WT using Pfx platinum DNA polymerase (Invitrogen). Primers are shown in the Table S3. The PCR products were digested with SacI-HF and Sall-HF and ligated with Quick T4-DNA ligase (New England BioLabs) into pBAD30 plasmid previously digested with the same enzymes.

Site-directed mutagenesis

Site-directed mutagenesis was performed following the Quik-Change protocol (Stratagene, Cedar Creek, TX) and with the *Pfu*Turbo (Stratagene) high-fidelity polymerase. For each reaction, 25–50 ng of pBADaafDA or pET101D-AggAdsA-O104H4 plasmid was combined with 10 pmol of each of the complementary primers. Reactions were carried out according to the manufacturer's protocol. Primers used to generate the single point mutations are shown in Table S3. All constructs were verified by Sanger sequencing at the University of Virginia DNA Core Facility or Sequencing service of Turku Centre for Biotechnology. EAEC042aafa and DH5 α were transformed with mutated pBADaafDA and pET101D-AggAdsA-O104H4, respectively, by heat shock at 42°C, rescued with SOC media, and selectively grown on LB-agar with kanamycin (pBADaafDA) or ampicillin (pET101D-AggAdsA-O104H4).

SDS-PAGE and immunoblotting

Protein analysis of AafA constructs harboring site mutations was performed by immunoblotting. EAEC042aafa transformed with pBADaafDA harboring site mutations was grown until OD₆₀₀ = 0.6, then induced with 2% arabinose until an OD₆₀₀ = 1.2. 1×10^7 cells were resuspended in Laemmli buffer, boiled and proteins were separated by 4–15% gradient acrylamide SDS-PAGE and transferred by 1 h at 100 V in Towbin's buffer onto nitrocellulose membranes (BioRad, Hercules CA). Membranes were blocked with 5% skim milk in phosphate-buffered saline/Tween, incubated with 200 ng/mL primary anti-AafA rabbit polyclonal antibody followed by 40 ng/mL of secondary HRP-conjugated anti-rabbit antibody (KPL, Gaithersburg, MD). Results were visualized directly on nitrocellulose membranes after exposure with TMB membrane peroxidase substrate (KPL, Gaithersburg, MD, USA).

Biofilm quantification

Biofilm formation was measured as previously described [12,13]. Briefly, bacteria were grown at 37°C overnight in LB in 24-well plates and induced with 0.2% arabinose. After washing the substratum with PBS and fixation with 10% (vol/vol) formalin for 10 minutes, bacteria were stained with 0.5% crystal violet for 5 minutes and solubilized with 70% ethanol for 5 minutes. The resulting solution was transferred to a microtiter plate where absorbance was read at 450 nm.

Bacterial binding to fibronectin

Quantification of bacterial binding to fibronectin was performed as previously described [23]. Briefly, wells of microtiter plates were coated with a solution of 25 μ g/ml of Fn protein in 100 mM Tris-HCl buffer, pH 8.0, overnight at 4°C. Plates were washed 5 times with phosphate-buffered saline (PBS) to remove unbound protein and blocked with 5% milk in PBS for 4 h at 4°C. Wells were then washed five times prior to the addition of the bacteria. 1 ml Dulbecco's modified Eagle's medium (DMEM)/0.5% glucose medium containing 1×10^8 bacteria at 37°C for 4 h were added to the wells. After the wells were washed 5 times with PBS, the bacterial cells that adhered to the wells were collected by

scraping them into PBS with 0.1% (vol/vol) Triton X-100; serial dilutions were plated onto LB agar plates supplemented with ampicillin. The number of adherent bacteria was determined by counting the resulting colonies in duplicate.

Immunofluorescence microscopy

Surface expression of AafA derivatives on EAEC042 was evaluated by indirect immunofluorescence assay (IFA). After induction of the bacterial strains with 0.2% arabinose, cells were harvested, washed twice with PBS, and incubated with a polyclonal anti-aafA antibody (2 μ g/ml) in PBS plus 1% bovine serum albumin for 1h at room temperature with agitation. Cells were washed twice with PBS and incubated with a goat anti-rabbit IgG-fluorescein isothiocyanate (FITC) conjugate (10 μ g/ml) for 30 min at room temperature in the dark. The washings were repeated, and the samples were resuspended in DAPI (4 mg/liter), and applied to slides for viewing by fluorescence microscopy.

Negative stain electron microscopy

2 μ l samples of sheared AAF/I fimbriae at ~ 100 μ g/ml and ~ 10 μ g/ml were applied to glow-discharged continuous-carbon-coated copper grids (Agar Scientific, UK), washed with 30 μ l 2% (w/v) uranyl acetate, blotted and air dried. Microscopy was performed using a Philips CM200 FEG electron microscope.

SPR binding assay

A Biacore X100 system (GE Healthcare) was used for all biosensor experiments. Fibronectin (Sigma) (approximately 1800 resonance units (RU)) was immobilized on flow cell 2 of a CM5 Sensor Chip by amine coupling using an Amine Coupling Kit (GE Healthcare). To record the association and dissociation curves, varying concentrations of subunit were injected into flow cell 2 of the chip for 3 min followed by flushing of the cell with 10 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% Tween 20 (HBS-EP) for 3 min at a flow rate of 10 ml min⁻¹. Identical samples were injected over a control flow cell to determine non-specific binding, which was subtracted from the experimental curves. The sensor chip was regenerated with 0.1% SDS. The equilibrium constants were determined by applying a one receptor binding model, using the Biacore X100 evaluation software and Smitfit/HLFIT program.

Statistical analysis

Statistical significance between means was analyzed using the unpaired Student's *t* test with a threshold *P* value of 0.05. Values are expressed as the means of three experiments with one standard deviation errors.

Supporting Information

Figure S1 The primary sequence alignment of major subunits from AAF and Afa/Dr families. The invariant and conserved positions are shaded in cyan and yellow, respectively. (TIFF)

Figure S2 Protein sequences of designed constructs. Blue, signal peptide; orange, His-tag; green, linker sequence; red, donor strand sequence. (TIFF)

Figure S3 Dynamic structure of AAF subunits. (A) NMR Relaxation properties of AafAdsA. Top panel: Black: Random Coil Index (RCI) predicted by TALOS+ based on AafA-dsc backbone chemical shifts. Region from residue number 125 to 135 shows high flexibility. Red: RMSD of each AafA residue generated

by ARIA/CNS. Middle panel: T_2 relaxation analysis of AafA-dsc. Bottom panel: ^1H - ^{15}N Heteronuclear NOE spectrum of AafAdsA. (B) Superimposition of the 10 best NMR structures of AafAdsA. (C) Structural superposition of the eight independently refined molecules from the asymmetric unit of the AggBdsA crystal. In A and C, the donor strand and linker sequence are shown in yellow and blue, respectively. The flexible N-terminal residues in the NMR structure are colored magenta. (TIFF)

Figure S4 Structural superposition of AggAdsA (magenta), AafAdsA (marine), and AfaEdsE (pdb: 2ixq) [35] (green) (stereo view) showing structurally variable segments. (TIFF)

Figure S5 Localization of key residues conserved across the major subunits of both AAF/I-III and AFA/Dr families (shaded in Fig. S1). These include the invariant Cys17, Cys49, Gly127 and Tyr129; the highly conserved Leu11 in the donor strand and Asp97, which is Asn in AAF/I from EAEC strain 17. Also shown are residues conserved within AAF/I-III major subunits only (i.e. not present in the AFA/Dr family), which include Lys55, Trp59, Gly60, Thr80 and Leu81 residues (from those shaded blue in Fig. 1). Side-chains are shown as balls on sticks on the structure of AggAdsA as a stereo view. Residues are numbered according to the sequence of AggA. (TIFF)

Figure S6 Modelling of AAF fibers (A) Molecular surface rendering of a model for AAF fibers. Fragments containing tip minor subunits and four major subunits are shown. Conserved surface residues (Fig. 4) are painted in red. (B) Structural superposition of the Caf1':Caf1" fragment of the crystal structure of the Caf1M:Caf1':Caf1" mini-fiber of the F1 antigen (PDB accession number 1Z9S) and solution structure of the AfaDdsE-AfaEdsE fusion protein representing the tip complex of Afa-III fimbriae (dsE, donor strand of AfaE, PDB accession number 2IXQ). The Caf1':Caf1" fragment is shown in magenta except the donor strand, which is shown in red. The AfaDdsE-AfaEdsE fusion is painted in green, except the dsE donor strand complementing the AfaD subunits (blue) and the linker sequence connecting dsE to AfaD (cyan). The N and C termini of protein chains are labeled. Note that the angle between adjacent subunits in the F1 and Afa-III fibers differs by $\sim 25^\circ$. (TIFF)

Figure S7 Fragment of the tip complex in AAF/I demonstrating the difference in the topology of disulfide bonds (balls on sticks) in the minor (AggB, green) and major (AggA, cyan) subunits (cartoon diagram, stereo view). Note that in AggA, the disulfide bond connects the α helix in the BC₁ loop with the donor strand linker, whereas in AggB, the disulfide bond connects the BC₁ loop with the end of strand F. (TIFF)

Figure S8 Biacore analysis of AggAdsA (A) and AggBdsA (B) binding to fibronectin. Left panels: SPR sensograms recorded for different concentration of AggAdsA or AggBdsA. Right panels: Saturation curves (one binding site model). (TIFF)

Figure S9 Expression of AafA harboring site mutations in EAEC. Protein analysis of AafA constructs harboring site

mutations was performed by immunoblotting. EAEC *aafA* mutant, was transformed with pBADaafDA harboring site mutations and grown until OD₆₀₀ = 0.6, then induced with 2% arabinose until an OD₆₀₀ = 1.2 was reached. 1×10^7 cells were resuspended in Laemmli buffer, boiled and proteins were separated by 4–15% gradient SDS-PAGE and transferred onto nitrocellulose membranes. Membranes were probed with an anti-AafA rabbit polyclonal antibody followed by HRP-conjugated anti-rabbit antibody. All residues were mutated to alanine except where noted. (TIFF)

Figure S10 Comparison of structure and stability of the wild type (WT) AggAdsA and AggAdsA carrying three mutations: Lys73Ala, Lys76Ala, and Lys78Ala. (A) Far UV circular dichroism (CD) spectra of the wild type AggAdsA at 20 (solid line) and 80°C (dashed line) and mutant AggAdsA at 20°C (dotted line). 20–40 g/l samples of the proteins in 10 mM Hepes, pH 7.4 and 50 mM NaCl were diluted to the concentration of 3 g/l with 20 mM phosphate buffer, pH 7.0, and CD spectra were recorded on a J-810 dichrograph (Jasco) equipped with a PTC-423S temperature control system in a 0.01 cm quartz cuvette. (B) Temperature stability of the wild type and mutant AggAdsA. Temperature dependence of the CD signal was recorded at a heating rate of 1°C min^{-1} at 218 nm for the wild type AggAdsA and at 230 nm for the mutant in a 0.2 cm cuvette. The protein concentration was 0.21 g/ml. The data were normalized and fitted to a two-state denaturation process. The transition temperatures for the wild type and mutant AggAdsA have similar values, 66.8 ± 0.4 and $65.4 \pm 0.2^\circ\text{C}$, respectively, suggesting that the introduced mutations did not affect protein stability. (TIFF)

Figure S11 Bovine serum albumin (BSA) does not affect the AggAdsA binding to fibronectin. SPR sensograms for the binding of AggAdsA and fibronectin were recorded at 40 μM concentration of AggAdsA in the absence (black solid line) and presence of 240 μM BSA (red dashed line). (TIFF)

Table S1 NMR structural constraints and structure statistics for AafAdsA. (PDF)

Table S2 Diffraction data and refinement statistics. (PDF)

Table S3 Primers used in the study. (PDF)

Acknowledgments

We would also like to thank the beamline scientists of ID23-1 at the ESRF and I24 at the DLS.

Author Contributions

Conceived and designed the experiments: AAB YY NP JAG IM JPN AVZ SM. Performed the experiments: AAB YY NP JAG WcL JPN SR MT BL KGI FRP IM. Analyzed the data: AAB YY NP JAG JM BL KGI FRP IM JPN AVZ SM. Contributed reagents/materials/analysis tools: EC JM IM JPN AVZ SM. Wrote the paper: AAB YY JAG EC KGI FRP IM JPN AVZ SM.

References

- Nataro JP, Kaper JB, Robins-Browne R, Prado V, Vial P, et al. (1987) Patterns of adherence of diarrheagenic Escherichia coli to HEp-2 cells. *Pediatr Infect Dis J* 6: 829–831.
- Okhuysen PC, Dupont HL (2010) Enterotoxigenic Escherichia coli (EAEC): a cause of acute and persistent diarrhea of worldwide importance. *J Infect Dis* 202: 503–505.

3. Wanke CA, Schorling JB, Barrett IJ, Desouza MA, Guerrant RL (1991) Potential role of adherence traits of *Escherichia coli* in persistent diarrhea in an urban Brazilian slum. *Pediatr Infect Dis J* 10: 746–751.
4. Mathewson JJ, Jiang ZD, Zumla A, Chintu C, Luo N, et al. (1995) HEP-2 cell-adherent *Escherichia coli* in patients with human immunodeficiency virus-associated diarrhea. *J Infect Dis* 171: 1636–1639.
5. Jiang ZD, Greenberg D, Nataro JP, Steffen R, DuPont HL (2002) Rate of occurrence and pathogenic effect of enteroaggregative *Escherichia coli* virulence factors in international travelers. *J Clin Microbiol* 40: 4185–4190.
6. Muniesa M, Hammerl JA, Hertwig S, Appel B, Brussow H (2012) Shiga toxin-producing *Escherichia coli* O104:H4: a new challenge for microbiology. *Appl Environ Microbiol* 78: 4065–4073.
7. Nataro JP (2011) Outbreak of hemolytic-uremic syndrome linked to Shiga toxin-producing enteroaggregative *Escherichia coli* O104:H4. *Pediatr Res* 70: 221.
8. Frank C, Werber D, Cramer JP, Askar M, Faber M, et al. (2011) Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *N Engl J Med* 365: 1771–1780.
9. Nataro JP, Yikang D, Giron JA, Savarino SJ, Kothary MH, et al. (1993) Aggregative adherence fimbria I expression in enteroaggregative *Escherichia coli* requires two unlinked plasmid regions. *Infect Immun* 61: 1126–1131.
10. Czczulin JR, Balepur S, Hicks S, Phillips A, Hall R, et al. (1997) Aggregative adherence fimbria II, a second fimbrial antigen mediating aggregative adherence in enteroaggregative *Escherichia coli*. *Infect Immun* 65: 4135–4145.
11. Bernier C, Gounon P, Le Bouguenec C (2002) Identification of an aggregative adherence fimbria (AAF) type III-encoding operon in enteroaggregative *Escherichia coli* as a sensitive probe for detecting the AAF-encoding operon family. *Infect Immun* 70: 4302–4311.
12. Boisen N, Struve C, Scheutz F, Krogfelt KA, Nataro JP (2008) New adhesin of enteroaggregative *Escherichia coli* related to the Afa/Dr/AAF family. *Infect Immun* 76: 3281–3292.
13. Sheikh J, Hicks S, Dall'Agnol M, Phillips AD, Nataro JP (2001) Roles for Fis and YafK in biofilm formation by enteroaggregative *Escherichia coli*. *Mol Microbiol* 41: 983–997.
14. Busch A, Waksman G (2012) Chaperone–usher pathways: diversity and pilus assembly mechanism. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367: 1112–1122.
15. Nuccio S-P, Bäuml AJ (2007) Evolution of the Chaperone/Usher Assembly Pathway: Fimbrial Classification Goes Greek. *Microbiology and Molecular Biology Reviews* 71: 551–575.
16. Zav'yalov V, Zav'yalov A, Zav'yalova G, Korpela T (2010) Adhesive organelles of Gram-negative pathogens assembled with the classical chaperone/usher machinery: structure and function from a clinical standpoint. *FEMS Microbiol Rev* 34: 317–378.
17. Zav'yalov A, Zav'yalova G, Korpela T, Zav'yalov V (2007) FGL chaperone-assembled fimbrial polyadhesins: anti-immune armament of Gram-negative bacterial pathogens. *FEMS Microbiol Rev* 31: 478–514.
18. Jouve M, Garcia MI, Courcoux P, Labigne A, Gounon P, et al. (1997) Adhesion to and invasion of HeLa cells by pathogenic *Escherichia coli* carrying the afa-3 gene cluster are mediated by the AfaE and AfaD proteins, respectively. *Infect Immun* 65: 4082–4089.
19. Korotkova N, Yarova-Yarovaya Y, Tcheshnokova V, Yazvenko N, Carl MA, et al. (2008) *Escherichia coli* DraE adhesin-associated bacterial internalization by epithelial cells is promoted independently by decay-accelerating factor and carcinoembryonic antigen-related cell adhesion molecule binding and does not require the DraD invasin. *Infect Immun* 76: 3869–3880.
20. Zalewska B, Piatek R, Bury K, Samet A, Nowicki B, et al. (2005) A surface-exposed DraD protein of uropathogenic *Escherichia coli* bearing Dr fimbriae may be expressed and secreted independently from DraC usher and DraE adhesin. *Microbiology* 151: 2477–2486.
21. Harrington SM, Strauman MC, Abe CM, Nataro JP (2005) Aggregative adherence fimbriae contribute to the inflammatory response of epithelial cells infected with enteroaggregative *Escherichia coli*. *Cell Microbiol* 7: 1565–1578.
22. Velarde JJ, Varney KM, Inman KG, Farfan M, Dudley E, et al. (2007) Solution structure of the novel dispersin protein of enteroaggregative *Escherichia coli*. *Mol Microbiol* 66: 1123–1135.
23. Farfan MJ, Inman KG, Nataro JP (2008) The major pilin subunit of the AAF/II fimbriae from enteroaggregative *Escherichia coli* mediates binding to extracellular matrix proteins. *Infect Immun* 76: 4378–4384.
24. Bernier C, Gounon P, Le Bouguenec C (2002) Identification of an Aggregative Adherence Fimbria (AAF) Type III-Encoding Operon in Enteroaggregative *Escherichia coli* as a Sensitive Probe for Detecting the AAF-Encoding Operon Family. *Infection and Immunity* 70: 4302–4311.
25. Zavialov AV, Berglund J, Pudney AF, Fooks LJ, Ibrahim TM, et al. (2003) Structure and biogenesis of the capsular F1 antigen from *Yersinia pestis*: preserved folding energy drives fiber formation. *Cell* 113: 587–596.
26. Anderson KL, Billington J, Pettigrew D, Cota E, Simpson P, et al. (2004b) An atomic resolution model for assembly, architecture, and function of the Dr adhesins. *Mol Cell* 15: 647–657.
27. Roy SP, Rahman MM, Yu XD, Tuittila M, Knight SD, et al. (2012) Crystal structure of enterotoxigenic *Escherichia coli* colonization factor CS6 reveals a novel type of functional assembly. *Mol Microbiol* 86: 1100–1115.
28. Bao R, Nair MKM, Tang W-k, Esser L, Sadhukhan A, et al. (2013) Structural basis for the specific recognition of dual receptors by the homopolymeric pH 6 antigen (Psa) fimbriae of *Yersinia pestis*. *Proceedings of the National Academy of Sciences* 110: 1065–1070.
29. Remaut H, Rose RJ, Hannan TJ, Hultgren SJ, Radford SE, et al. (2006) Donor-Strand Exchange in Chaperone-Assisted Pilus Assembly Proceeds through a Concerted β Strand Displacement Mechanism. *Molecular Cell* 22: 831–842.
30. Sauer FG, Pinkner JS, Waksman G, Hultgren SJ (2002) Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation. *Cell* 111: 543–551.
31. Barnhart MM, Pinkner JS, Soto GE, Sauer FG, Langermann S, et al. (2000) PapD-like chaperones provide the missing information for folding of pilin proteins. *Proc Natl Acad Sci U S A* 97: 7709–7714.
32. Zavialov AV, Tischenko VM, Fooks LJ, Brandsdal BO, Aqvist J, et al. (2005) Resolving the energy paradox of chaperone/usher-mediated fibre assembly. *Biochem J* 389: 685–694.
33. Piatek R, Bruzdziak P, Wojciechowski M, Zalewska-Piatek B, Kur J (2010) The noncanonical disulfide bond as the important stabilizing element of the immunoglobulin fold of the Dr fimbrial DraE subunit. *Biochemistry* 49: 1460–1468.
34. Crespo MD, Puorger C, Schärer MA, Eidam O, Grütter MG, et al. (2012) Quality control of disulfide bond formation in pilus subunits by the chaperone FimC. *Nat Chem Biol* 8: 707–713.
35. Cota E, Jones C, Simpson P, Altroff H, Anderson KL, et al. (2006) The solution structure of the invasive tip complex from Afa/Dr fibrils. *Mol Microbiol* 62: 356–366.
36. Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Research* 38: W545–549.
37. Sauer FG, Remaut H, Hultgren SJ, Waksman G (2004) Fiber assembly by the chaperone–usher pathway. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1694: 259–267.
38. Berjanskii MV, Wishart DS (2008) Application of the random coil index to studying protein flexibility. *J Biomol NMR* 40: 31–48.
39. Lawrence MC, Colman PM (1993) Shape complementarity at protein/protein interfaces. *J Mol Biol* 234: 946–950.
40. Garcia M-I, Gounon P, Courcoux P, Labigne A, Le Bouguenec C (1996) The afimbrial adhesive sheath encoded by the afa-3 gene cluster of pathogenic *Escherichia coli* is composed of two adhesins. *Molecular Microbiology* 19: 683–693.
41. Garcia M-I, Jouve M, Nataro JP, Gounon P, Le Bouguenec C (2000) Characterization of the AfaD-like family of invasins encoded by pathogenic *Escherichia coli* associated with intestinal and extra-intestinal infections. *FEBS Letters* 479: 111–117.
42. Salih O, Remaut H, Waksman G, Orlova EV (2008) Structural Analysis of the Saf Pilus by Electron Microscopy and Image Processing. *Journal of Molecular Biology* 379: 174–187.
43. Korotkova N, Yang Y, Le Trong I, Cota E, Demeler B, et al. (2008) Binding of Dr adhesins of *Escherichia coli* to carcinoembryonic antigen triggers receptor dissociation. *Molecular Microbiology* 67: 420–434.
44. Das M, Hart-Van Tassel A, Urvil PT, Lea S, Pettigrew D, et al. (2005) Hydrophilic Domain II of *Escherichia coli* Dr Fimbriae Facilitates Cell Invasion. *Infection and Immunity* 73: 6119–6126.
45. Boll EJ, Struve C, Sander A, Demma Z, Nataro JP, et al. (2012) The fimbriae of enteroaggregative *Escherichia coli* induce epithelial inflammation in vitro and in a human intestinal xenograft model. *J Infect Dis* 206: 714–722.
46. Strauman MC, Harper JM, Harrington SM, Boll EJ, Nataro JP (2010) Enteroaggregative *Escherichia coli* disrupts epithelial cell tight junctions. *Infect Immun* 78: 4958–4964.
47. de Oliveira-Garcia D, Dall'Agnol M, Rosales M, Azzuz AC, Alcantara N, et al. (2003) Fimbriae and adherence of *Stenotrophomonas maltophilia* to epithelial cells and to abiotic surfaces. *Cell Microbiol* 5: 625–636.
48. Forest KT, Dunham SA, Koomey M, Tainer JA (1999) Crystallographic structure reveals phosphorylated pilin from *Neisseria*: phosphoserine sites modify type IV pilus surface chemistry and fibre morphology. *Molecular Microbiology* 31: 743–752.
49. Needleman DJ, Ojeda-Lopez MA, Raviv U, Miller HP, Wilson L, et al. (2004) Higher-order assembly of microtubules by counterions: From hexagonal bundles to living necklaces. *Proceedings of the National Academy of Sciences of the United States of America* 101: 16099–16103.
50. Bingham RJ, Rudiño-Piñera E, Meenan NAG, Schwarz-Linek U, Turkenburg JP, et al. (2008) Crystal structures of fibronectin-binding sites from *Staphylococcus aureus* FnBPA in complex with fibronectin domains. *Proceedings of the National Academy of Sciences* 105: 12254–12258.
51. Konar M, Sachin O, Priya A, Ghosh S (2012) Identification of key proteins of cultured human intestinal cells involved in interaction with enteroaggregative *Escherichia coli*. *FEMS Immunology & Medical Microbiology* 66: 177–190.
52. Izquierdo M, Navarro-García F, Nava-Acosta R, Nataro JP, Ruiz-Perez F, et al. (2014) Identification of cell surface-exposed proteins involved in the fimbria-mediated adherence of enteroaggregative *Escherichia coli* to intestinal cells. *Infect Immun* 82: 1719–1724.
53. Nuccitelli A, Cozzi R, Gourlay LJ, Donnarumma D, Necchi F, et al. (2011) Structure-based approach to rationally design a chimeric protein for an effective vaccine against Group B Streptococcus infections. *Proceedings of the National Academy of Sciences* 108: 10278–10283.
54. Zhang C, Zhang W (2010) *Escherichia coli* K88ac Fimbriae Expressing Heat-Labile and Heat-Stable (StA) Toxin Epitopes Elicit Antibodies That Neutralize

- Cholera Toxin and STa Toxin and Inhibit Adherence of K88ac Fimbrial *E. coli*. *Clinical and Vaccine Immunology* 17: 1859–1867.
55. Huesca M, Sun Q, Peralta R, Shivji GM, Sauder DN, et al. (2000) Synthetic Peptide Immunogens Elicit Polyclonal and Monoclonal Antibodies Specific for Linear Epitopes in the D Motifs of *Staphylococcus aureus* Fibronectin-Binding Protein, Which Are Composed of Amino Acids That Are Essential for Fibronectin Binding. *Infection and Immunity* 68: 1156–1163.
 56. Dormitzer PR, Grandi G, Rappuoli R (2012) Structural vaccinology starts to deliver. *Nat Rev Micro* 10: 807–813.
 57. Pakharukova N, Tuittila M, Zavialov A (2013) Crystallization and sulfur SAD phasing of AggA, the major subunit of aggregative adherence fimbriae type I from the *Escherichia coli* strain that caused an outbreak of haemolytic-uraemic syndrome in Germany. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 69: 1389–1392.
 58. Yang Y, Garnett JA, Matthews S (2011) Crystallization and initial crystallographic analysis of AafA: the major adhesive subunit of the enteroaggregative *Escherichia coli* AAF/II pilus. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 67: 454–456.
 59. Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* 66: 125–132.
 60. Jedrzejczak R, Dauter Z, Dauter M, Piatek R, Zalewska B, et al. (2006) Structure of DraD invasins from uropathogenic *Escherichia coli*: a dimer with swapped [beta]-tails. *Acta Crystallographica Section D* 62: 157–164.
 61. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, et al. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58: 1948–1954.
 62. Winn MD, Isupov MN, Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallographica Section D-Biological Crystallography* 57: 122–133.
 63. Marchant J, Sawmynaden K, Saouros S, Simpson P, Matthews S (2008) Complete resonance assignment of the first and second apple domains of MIC4 from *Toxoplasma gondii*, using a new NMRView-based assignment aid. *Biomolecular Nmr Assignments* 2: 119–121.
 64. Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, et al. (2007) ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* 23: 381–382.
 65. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS plus: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular Nmr* 44: 213–223.