


RESEARCH ARTICLE

Ucl fimbriae regulation and glycan receptor specificity contribute to gut colonisation by extra-intestinal pathogenic *Escherichia coli*

Steven J. Hancock^{1,2}[✉], Alvin W. Lo^{1,2}[✉], Thomas Ve^{1,3}[✉], Christopher J. Day³[✉], Lendl Tan^{1,2}, Alejandra A. Mendez⁴, Minh-Duy Phan^{1,2}, Nguyen Thi Khanh Nhu^{1,2}, Kate M. Peters^{1,2}, Amanda C. Richards⁴, Brittany A. Fleming⁴, Chyden Chang^{1,2}, Dalton H. Y. Ngu^{1,2}, Brian M. Forde^{1,2}, Thomas Haselhorst³, Kelvin G. K. Goh^{1,2}, Scott A. Beatson^{1,2}, Michael P. Jennings³, Matthew A. Mulvey⁴, Bostjan Kobe^{1,2,5}^{*}, Mark A. Schembri^{1,2*}

1 School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland, Australia, **2** Australian Infectious Diseases Research Centre, The University of Queensland, Brisbane, Queensland, Australia, **3** Institute for Glycomics, Griffith University Gold Coast Campus, Gold Coast, Queensland, Australia, **4** Division of Microbiology and Immunology, Department of Pathology, University of Utah School of Medicine, Salt Lake City, Utah, United States of America, **5** Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia

 These authors contributed equally to this work.

 Current address: Wellcome-Wolfson Institute for Experimental Medicine, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, United Kingdom.

* m.schembri@uq.edu.au


 OPEN ACCESS

Citation: Hancock SJ, Lo AW, Ve T, Day CJ, Tan L, Mendez AA, et al. (2022) Ucl fimbriae regulation and glycan receptor specificity contribute to gut colonisation by extra-intestinal pathogenic *Escherichia coli*. PLoS Pathog 18(6): e1010582. <https://doi.org/10.1371/journal.ppat.1010582>

Editor: Olivera Francetic, Institut Pasteur, FRANCE

Received: November 2, 2021

Accepted: May 9, 2022

Published: June 14, 2022

Copyright: © 2022 Hancock et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: This work was supported by National Health and Medical Research Council (NHMRC) grants 1106930, 1181958 and 2001431 to MAS; and 1071659, 1110971 and 1180826 to BK and 1108124 to MPJ and CJD. Additional support was provided by US National Institutes of Health grants AI135918 and GM134331 to MAM, a Chevron Undergraduate Research Scholarship for women in STEM to AAM, an NIH Microbial Pathogenesis T32

Abstract

Extra-intestinal pathogenic *Escherichia coli* (ExPEC) belong to a critical priority group of antibiotic resistant pathogens. ExPEC establish gut reservoirs that seed infection of the urinary tract and bloodstream, but the mechanisms of gut colonisation remain to be properly understood. Ucl fimbriae are attachment organelles that facilitate ExPEC adherence. Here, we investigated cellular receptors for Ucl fimbriae and Ucl expression to define molecular mechanisms of Ucl-mediated ExPEC colonisation of the gut. We demonstrate differential expression of Ucl fimbriae in ExPEC sequence types associated with disseminated infection. Genome editing of strains from two common sequence types, F11 (ST127) and UTI89 (ST95), identified a single nucleotide polymorphism in the *ucl* promoter that changes fimbriae expression via activation by the global stress-response regulator OxyR, leading to altered gut colonisation. Structure-function analysis of the Ucl fimbriae tip-adhesin (UclD) identified high-affinity glycan receptor targets, with highest affinity for sialyllacto-N-fucopentose VI, a structure likely to be expressed on the gut epithelium. Comparison of the UclD adhesin to the homologous UcaD tip-adhesin from *Proteus mirabilis* revealed that although they possess a similar tertiary structure, apart from lacto-N-fucopentose VI that bound to both adhesins at low-micromolar affinity, they recognize different fucose- and glucose-containing oligosaccharides. Competitive surface plasmon resonance analysis together with co-structural investigation of UcaD in complex with monosaccharides revealed a broad-specificity glycan binding pocket shared between UcaD and UclD that could accommodate these interactions. Overall, our study describes a mechanism of adaptation that augments

Training Grant AI055434 to ACR and an Australian Research Council (ARC) Discovery Project to MPJ (DP170104692). BK is an ARC Laureate Fellow (FL180100109). TV is an ARC Future Fellow (FT200100572), received ARC DECRA funding (DE170100783) and is also an NHMRC Investigator (1196590). MPJ is an NHMRC Principal Research Fellow (1138466). 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.

establishment of an ExPEC gut reservoir to seed disseminated infections, providing a pathway for the development of targeted anti-adhesion therapeutics.

Author summary

ExPEC infection of the urinary tract and bloodstream is frequently seeded from an intestinal reservoir, necessitating an understanding of the mechanisms that promote gut colonisation. Here we employed molecular and structural approaches to define the regulation and function of ExPEC Ucl fimbriae as a gut colonisation factor. We describe how mutations in the non-coding regulatory region of the *ucl* promoter cause increased Ucl fimbriae expression and promote enhanced gut colonisation via tuned induction by a global regulator that senses oxygen stress. We further define the glycan receptor targets of Ucl fimbriae and characterise the structural features of the Ucl adhesin that facilitate these interactions. These findings explain how ExPEC can adapt to survival in the gut to seed extra-intestinal infection.

Introduction

Extra-intestinal pathogenic *Escherichia coli* (ExPEC) are a major cause of multiple severe human infections, including infections of the urinary tract, prostate, bloodstream and meninges [1]. ExPEC is also a lead driver of current, new and emerging antibiotic resistance [2,3], underscored by its top-three ranking in the WHO list of 12 antibiotic resistant priority pathogens. The primary pathway of human ExPEC infection involves initial colonisation of the gut, leading to the formation of a reservoir that can seed dissemination to extra-intestinal sites. An enhanced capacity to colonise the gut may contribute to the success of globally disseminated ExPEC clones that can be differentiated based on their multi-locus sequence type (ST), the most common being ST131 [3,4].

Adherence to host tissue epithelial surfaces represents a critical step in ExPEC pathogenesis and disease development. ExPEC adherence is primarily mediated by fimbriae, long hair-like polymeric surface structures that facilitate colonisation by specific interaction with target receptors. The assembly of ExPEC fimbriae engages a conserved chaperone-usher (CU) pathway [5]. CU fimbriae biogenesis involves a dedicated periplasmic chaperone and outer membrane-bound usher, which serve as an assembly apparatus for the growing organelle. Individual fimbriae are typically comprised of ~1000 major subunit proteins polymerized into a right-handed helical structure, and a receptor-binding adhesin located at the distal tip of the organelle [6]. Tip adhesins are two-domain proteins; they possess a C-terminal pilin domain that connects the adhesin to the main fimbrial shaft, and an N-terminal lectin domain that mediates attachment to specific ligands, therefore dictating host range and tissue tropism [5].

Most ExPEC encode the capacity to produce multiple different types of CU fimbriae, the genes for which often lie within large genomic islands (GIs). For example, the reference strain UTI89 contains nine different CU fimbriae loci [7]. Type 1 and P fimbriae are the best-characterized CU fimbriae and contribute to colonisation of the bladder and kidney, respectively [8]. Ucl fimbriae (also previously annotated as F17-like or Uca-like fimbriae) are another type of CU fimbriae that contribute to ExPEC infection [9]. The *ucl* fimbrial operon consists of four genes (*uclABCD*), which encode a major subunit protein (UclA), chaperone (UclB), usher (UclC) and adhesin (UclD), respectively. Ucl fimbriae promote strong biofilm formation and

attachment to exfoliated human uroepithelial cells [10,11]. Mutation of the *ucl* fimbrial genes in the UPEC strain UTI89 impacts intestinal colonization via interaction of the UclD adhesin with undefined O-linked oligosaccharides on colonic epithelial cells [12].

Ucl fimbriae are phylogenetically related to Uca fimbriae from *Proteus mirabilis* and F17/G fimbriae from enterotoxigenic *E. coli* (ETEC) [10]. *P. mirabilis* Uca fimbriae mediate adherence to desquamated uroepithelial cells and facilitate colonisation of the bladder and kidney in experimental mice [13,14]. F17/G fimbriae of ETEC, on the other hand, promote binding to N-acetyl- β -D-glucosamine (GlcNAc)-containing receptors on the microvilli of the intestinal epithelium of ruminants [15]. Although the binding of ExPEC Ucl fimbriae and *P. mirabilis* Uca fimbriae to a range of epithelial cells has been demonstrated, the precise receptor targets of these two closely related fimbriae remain unknown. In this study, we observed variation in the expression of Ucl fimbriae between different ExPEC strains, and identified a single nucleotide polymorphism in the *ucl* promoter region that affects Ucl fimbrial expression via differential activation by the global stress response regulator OxyR, which in turn alters intestinal colonisation. To define the receptor specificity of Ucl fimbriae, we determined the crystal structure of the UclD lectin binding domain (UclD^{LD}) and the lectin binding domain of the related UcaD adhesin (UcaD^{LD}). We show that both UclD^{LD} and UcaD^{LD} adopt a conserved tertiary structure, but bind to different glycans. Through crystallographic studies of UcaD^{LD} in complex with monosaccharides, as well as competitive surface plasmon resonance (SPR) analysis and molecular docking, we identified a broad-specificity binding pocket that can serve as an anchor point for receptor binding by interacting with fucose and glucose residues, which are common moieties in the identified glycan receptors. Together, these findings explain how ExPEC can adapt to survival in the gut and establish a persistent reservoir with the potential to seed extra-intestinal infection.

Results

Differential expression of the UclA major fimbrial subunit protein

We began by comparing the expression of UclA in the ExPEC reference strains F11 (ST127), UTI89 (ST95), and two strains from the globally disseminated multidrug resistant ST131 lineage (S77EC and HVM1299) [16, 17]. These strains are part of a restricted set of ExPEC STs that we showed possess the Ucl fimbrial genes (Fig A in [S1 Text](#)), and all possess the most dominant *uclA*-10 allelic variant, one of 15 *uclA* allelic variants that we identified in our dataset of 8,247 *E. coli* genomes (Fig B in [S1 Text](#)). Western blot analysis using a UclA-specific antibody revealed that all strains express UclA, and that UclA expression in F11 is significantly increased compared to the other three strains (Fig 1A). In both F11 and UTI89, the two ExPEC strains used for the remainder of this study, the *uclABCD* genes reside in a conserved GI located at the *leuX* tRNA (GI-*leuX*; Fig C in [S1 Text](#)), determined based on comparative analysis of the complete F11 genome (comprising a 5,048,308 base-pair chromosome and a single 114,223 base-pair F plasmid; [S1 Data](#)) and the UTI89 genome [7].

A single nucleotide change in the F11 *ucl* promoter region leads to increased transcription and expression of Ucl fimbriae

To understand the difference in UclA expression between F11 and UTI89, we first mapped the *ucl* transcriptional start site in F11 using 5' RACE (indicated together with promoter elements in Fig D in [S1 Text](#); source data in Fig E in [S1 Text](#)). We then aligned the *ucl* promoter region from F11, UTI89, S77EC and HVM1299. The *ucl* promoter region is conserved, with the exception of one change at a position 78 nucleotides upstream of the *uclA* transcription start site in F11 that differs from the other three strains (F11 T>G other strains). To examine if this

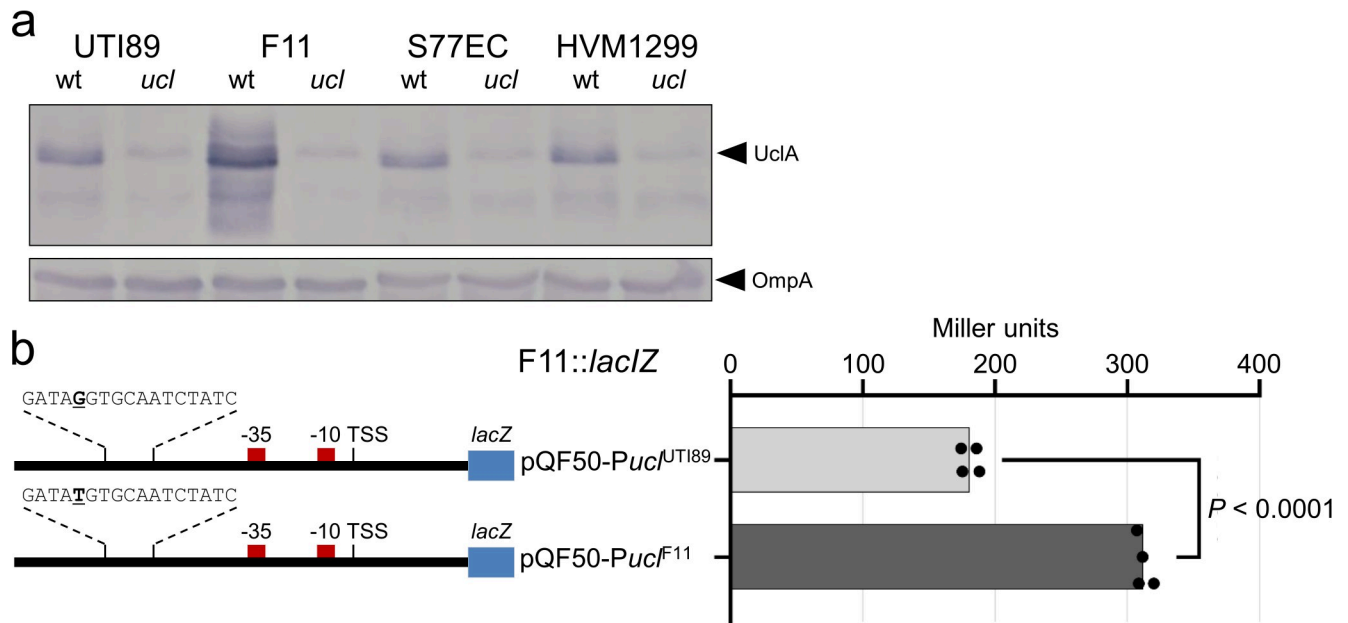


Fig 1. Expression of UclA differs among ExPEC strains. a) Whole-cell lysate western blot analysis of UTI89, F11, S77EC and HVM1299 and their respective *ucl* deletion mutants, using a polyclonal antibody generated against UclA^{F11} (18.9 kDa) and OmpA (loading control). b) Left, schematic diagram of the *ucl* promoter region from F11 and UTI89 cloned in the reporter plasmid pQF50. Indicated are the TSS, -10 and -35 promoter elements, and OxyR binding site. Right, β -galactosidase activity (measured in Miller units) for the pQF50-*PucI*^{F11}-*lacZ* versus pQF50-*PucI*^{UTI89}-*lacZ* fusion constructs in F11lacZ ($p < 0.0001$; one-way ANOVA with Sidak's multiple comparisons test).

<https://doi.org/10.1371/journal.ppat.1010582.g001>

single nucleotide change in F11 accounted for increased expression of UclA, the *ucl* promoter regions from F11 and UTI89 were PCR-amplified and cloned into the *lacZ* reporter plasmid pQF50, to generate plasmids pQF50-*PucI*^{F11}-*lacZ* and pQF50-*PucI*^{UTI89}-*lacZ*, respectively. The resultant plasmids were then used to transform F11 and UTI89 *lac*⁻ derivative strains (F11*lacI*-Z and UTI89*lacI*-Z, respectively), and β -galactosidase activity was measured following growth in Luria-Bertani (LB) broth to assess *ucl* promoter activity. Consistent with the western-blot analysis, β -galactosidase activity produced from plasmid pQF50-*PucI*^{F11}-*lacZ* was significantly higher than the corresponding activity from plasmid pQF50-*PucI*^{UTI89}-*lacZ* in both F11*lacI*-Z (Fig 1B) and UTI89*lacI*-Z (Fig D in S1 Text).

To further analyse the functional role of the F11 T>G UTI89 nucleotide in the *ucl* promoter region, we swapped the nucleotide on the chromosome of both strains using a recently described genome editing method [18]. Using this approach, we successfully exchanged the 'T⁽⁻⁷⁸⁾' nucleotide in the *ucl* promoter of F11 (to 'G') and the 'G⁽⁻⁷⁸⁾' nucleotide in the *ucl* promoter of UTI89 (to 'T'), to generate the strains F11-*PucI*^{T-78G} and UTI89-*PucI*^{G-78T}, respectively. Whole-genome Illumina sequencing confirmed that the F11-*PucI*^{T-78G} and UTI89-*PucI*^{G-78T} strains only contained the introduced single-nucleotide change. UclA-specific western-blot analysis of whole-cell lysates prepared from both sets of isogenic strains revealed that UclA expression was decreased in F11^{T-78G} compared to wild-type F11. Conversely, UclA expression was increased in UTI89^{G-78T} compared to wild-type UTI89 (Fig F in S1 Text). We also examined surface expression of Ucl fimbriae by whole-cell ELISA with our UclA-specific antibody, which reproduced the Ucl fimbriae expression profile in wild-type F11 versus F11-*PucI*^{T-78G} and wild-type UTI89 versus UTI89-*PucI*^{G-78T} strains (Fig G in S1 Text). Finally, transmission electron microscopy (TEM) in combination with immunogold labelling employing our UclA-specific antibody was used to directly examine the level of Ucl fimbriae produced by wild-type F11, F11-*PucI*^{T-78G} and F11 Δ *uclA*. Gold-particle labelling was

assessed from 100 randomly selected cells; 79% of wild-type F11 cells produced Ucl fimbriae compared to 12% of F11-*Pucl*^{T-78G} cells, while no labelling was observed for F11Δ*uclA* (Fig G in S1 Text). Taken together, our results demonstrate the single-nucleotide change to a 'T' at a position 78 nucleotides upstream of the *uclA* transcription start site in F11 is responsible for the enhanced expression of Ucl fimbriae in F11 compared to UTI89.

OxyR regulates Ucl fimbriae expression

To better understand the regulation at *Pucl*, we performed *in silico* DNA-binding protein predictions using the Virtual Footprint webserver for both the F11-*Pucl*^{78T} and the UTI89-*Pucl*^{78G} sequences [19]. The output showed high potential binding affinity of OxyR to both *Pucl*^{78T} (score = 13.57) and *Pucl*^{78G} (score = 14.38), noting this SNP occurs at position 5 of the previously defined OxyR consensus binding motif, and the *Pucl*^{78G} sequence is a closer match to this consensus sequence [20] (Figs 1B and 2A). Based on this result, we mutated the *oxyR* gene in both F11 and F11-*Pucl*^{T-78G} and analysed UclA expression by western blotting. We observed a drastic reduction in UclA expression for both F11*oxyR* and F11-*Pucl*^{T-78G}*oxyR* mutants, while complementation with the expression vector pBAD-OxyR restored UclA expression (Fig 2B). We also inserted a *lacZ* reporter as a transcriptional fusion to the *Pucl* promoter on the chromosome of F11*lacZ* and F11*lacZ*-*Pucl*^{T-78G}, and transformed both strains with a plasmid containing the *oxyR* gene (pOxyR). Beta-galactosidase activity of both strains increased with increasing concentration of the inducer arabinose (Fig 2C). However, *Pucl*_{F11} was activated to a greater degree than *Pucl*^{T-78G} at all concentrations of the inducer, revealing that OxyR is required for the expression of Ucl in a dose-dependent manner and the T-78G SNP reduces OxyR mediated activation of *Pucl*.

To experimentally demonstrate that OxyR binds directly to *Pucl*^{78G} and *Pucl*^{78T}, we expressed and purified a C-terminally 6xHis-tagged recombinant OxyR protein under non-denaturing conditions (Fig H in S1 Text). PCR-amplified Cy3-labelled DNA, representing a 261 bp region spanning the *Pucl*^{78G}/*Pucl*^{78T} promoter (-179 to +82 relative to the TSS), was incubated with increasing concentrations of OxyR-6xHis and visualised by gel electrophoresis. Mobility of both promoter elements was retarded by the addition of OxyR, with stronger affinity observed for *Pucl*^{78G} (first evidence of shift at 0.29 μM OxyR) compared to *Pucl*^{78T} (first evidence of shift at 0.57 μM OxyR) (Fig H in S1 Text). The addition of excess unlabelled competitor *Pucl*^{78G}/*Pucl*^{78T} DNA reversed binding of Cy3-labelled target DNA, demonstrating the specificity of this interaction (Fig H in S1 Text). However, binding of Cy3-labelled *Pucl*^{78G} DNA was less effectively reversed compared to Cy3-labelled *Pucl*^{78T} DNA, suggesting that OxyR binds to *Pucl*^{78G} with greater affinity than *Pucl*^{78T} in this *in vitro* binding assay. This is consistent with the *Pucl*^{78G} sequence being a closer match to the consensus OxyR binding site, and suggests there are additional regulatory cofactors involved in the activation of Ucl fimbriae expression. Finally, to assess the broader significance of SNPs in the OxyR binding site upstream of the *uclA* gene, we compared the sequence in 698 *ucl*-positive ST127 strains downloaded from Enterobase (noting F11 belongs to ST127). In total, 217/698 (31%) of ST127 strains contained SNPs in this site, with the F11-*Pucl*^{T-78G} SNP the most common (Fig I in S1 Text). Additional examination of two *ucl*-positive strains that we previously identified from the ECOR reference collection [10, 11] revealed that one strain, ECOR-60, expressed increased UclA and possessed a *Pucl*^{T-76C} SNP (Fig J in S1 Text), demonstrating that tuning of Ucl fimbriae expression can be caused by different SNPs in the OxyR binding site.

Increased Ucl fimbriae expression enhances colonisation of the mouse gut

Previous work has demonstrated that mutation of the Ucl fimbriae genes in UTI89 attenuates colonisation of the mouse gut [12]. Thus, we hypothesized that wild-type F11 would have a

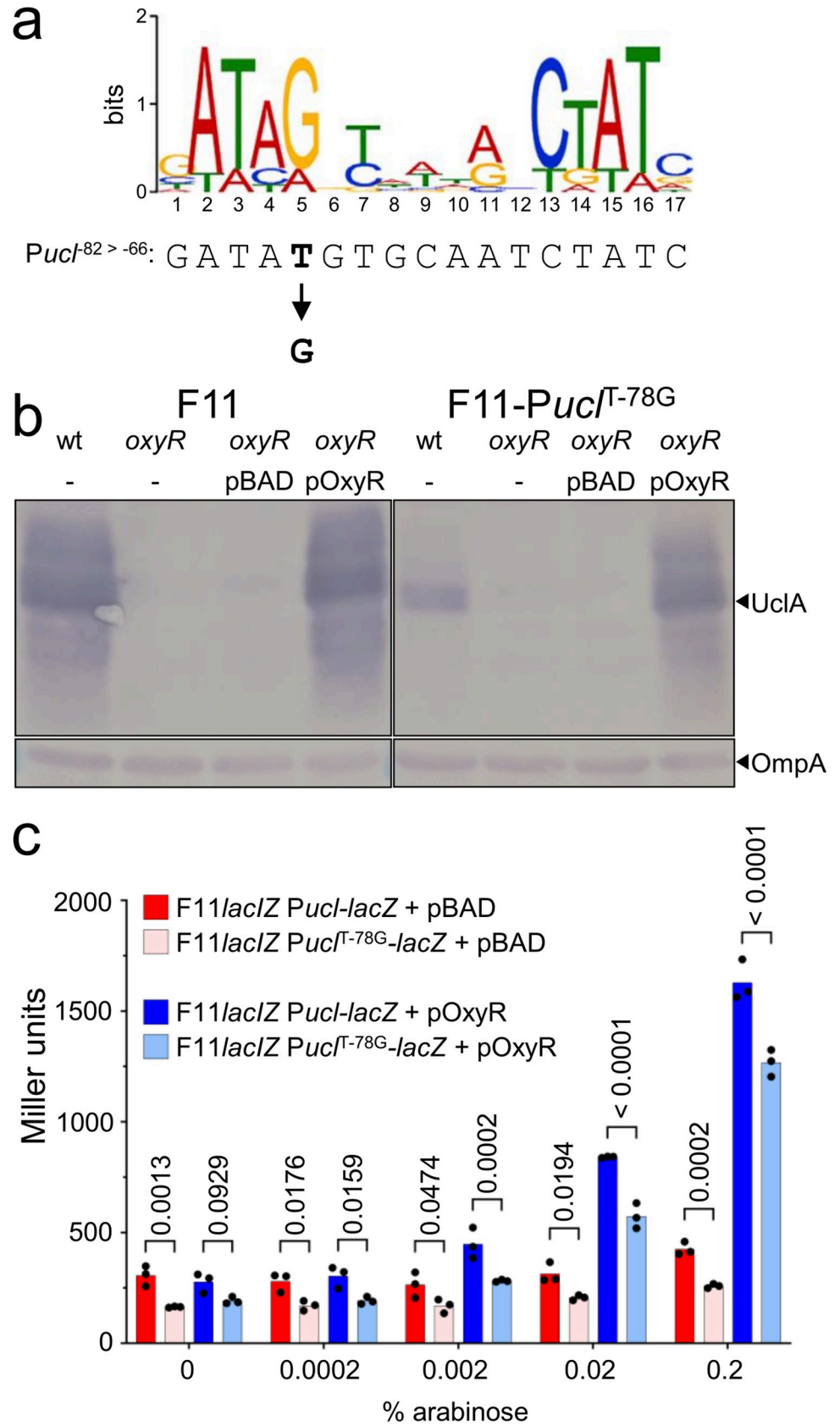


Fig 2. OxyR activates Ucl expression. a) Alignment of OxyR consensus binding motif [20] with the F11 *Pucl* positions -82 to -66 relative to the TSS indicated. Also indicated is the 'T' nucleotide at position -78, which is a 'G' nucleotide in UTI89 and the ST131 S77EC and HVM1299 strains. b) Whole-cell lysate western-blot analysis of F11 and F11^{T-78G} in the presence or absence of OxyR. c) Beta-galactosidase activity of F11*lacZ Pucl-lacZ* and F11*lacZ Pucl^{T-78G}-lacZ* harbouring pBAD (vector control; red and pink shading) or pOxyR (dark blue and light blue shading). Data points show the results of triplicate biological experiments, each representative of four technical replicates, and the bar indicates the mean. P-values calculated using two-way ANOVA and Sidak's multiple comparisons test on log₁₀ transformed data.

<https://doi.org/10.1371/journal.ppat.1010582.g002>

colonisation advantage over the F11-*Pucl^{T-78G}* mutant in similar experiments. First, we performed single infection experiments, where there was no significant difference in colonisation of the gut by wild-type F11 compared to F11-*Pucl^{T-78G}* and an F11Δ*ucl* mutant (Fig K in S1 Text). Next, we performed a more sensitive competitive experiment where mice were given a 1:1 mixture of two differentially tagged F11 strains (control) or F11 versus F11-*Pucl^{T-78G}* by oral gavage, and colonisation was monitored over time, taking advantage of kanamycin/chloramphenicol resistance markers. While there was no difference in colonisation by the differentially tagged control strains, wild-type F11 significantly outcompeted the F11-*Pucl^{T-78G}* mutant from days 7–15 post-infection (Fig 3). Thus, increased expression of Ucl fimbriae promotes enhanced colonisation of the mouse gut by F11 in a mixed infection experiment.

Receptor specificity of the UclD tip adhesin

To extend our investigation of Ucl fimbriae beyond its regulation, we performed a detailed structure-function analysis to characterise the binding specificity of Ucl fimbriae. We initially demonstrated the tip-location of the UclD adhesin using a recombinant strain expressing Ucl fimbriae (via plasmid pUcl^{F11}; MS7882) using double immuno-gold electron microscopy in combination with UclA and UclD specific antibodies (Fig L in S1 Text). Next, we determined the crystal structure of UclD^{LD} from F11 at 2.2 Å resolution (Table A in S1 Text). Given the

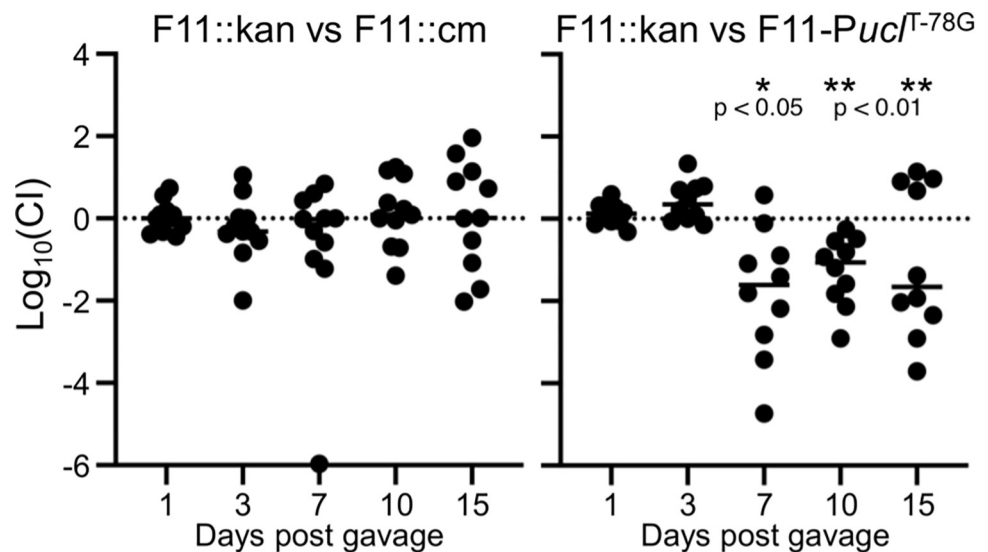


Fig 3. Increased Ucl fimbriae expression enhances mouse gut colonisation. Mice were inoculated with a 1:1 mixture of F11::kan vs F11::cm (control) or F11::kan vs F11-*Pucl^{T-78G}* (cm resistant). Each group contained 10–11 mice infected and monitored during two independent experiments. Differential resistance markers were used to determine bacterial loads and a competitive index was calculated for each mouse at each time point. Wilcoxon Signed Rank tests were performed on each group for each time point; * $P < 0.05$, ** $P < 0.01$.

<https://doi.org/10.1371/journal.ppat.1010582.g003>

close phylogenetic relationship between the Ucl fimbriae and the Uca fimbriae from *P. mirabilis* [10], in parallel we also determined the crystal structure of UcaD^{LD} (84% amino acid identity to UclD^{LD}) at 1.6 Å resolution. The *ucaABCD* genes are found in ~65% of *P. mirabilis* genomes and are also present in other *Proteus* species (Fig M in S1 Text), and the UcaD, UclD and the related GafD adhesin from F17/G fimbriae all share significant overall amino acid identity (Fig N in S1 Text). At the time of collecting our data, there was no structural model with sufficient sequence identity available in the PDB. Thus, we used iodide SAD phasing, using a sodium iodide derivative of UclD^{LD} that diffracted to 2.85 Å resolution, to obtain phase information. The model was then used as a template for solving the structure of UcaD^{LD} by molecular replacement. The final UclD^{LD} and UcaD^{LD} models contain residues 21–215 and 21–217, respectively. Electron density was not observed for residues 43–48 (UclD^{LD}) and 44–47 (UcaD^{LD}), suggesting that these regions have a disordered or flexible conformation in the crystals. As expected from the high sequence identity, the two structures are similar, with r.m.s.d. of 0.6 Å for 189 equivalent C α atoms. The UclD^{LD} and UcaD^{LD} structures are also very similar to the recently published UclD^{LD} structure from UTI89 (PDB: 5NWP) (r.m.s.d. of 0.2 Å for 189 equivalent C α atoms), and the UcaD^{LD} structure from *P. mirabilis* strain HI4320 (PDB: 6H2L) (r.m.s.d. of 0.6 Å for 189 equivalent C α atoms) [12, 21], respectively, as expected from their completely conserved protein sequences.

Both UclD and UcaD adopt a compact elongated Ig-like β -sandwich structure with a Greek-key topology, similar to other adhesin lectin domains (Fig 4A). A search for structures similar to UclD and UcaD in the PDB using the DALI server [22] revealed significant similarity to F17G like adhesins (Z-score of 14.3–15.8 when using UcaD^{LD} as the search model), although the sequence identity is only 14–16%. UclD^{LD} and UcaD^{LD} have a completely conserved pocket in the same region of the structure as the N-acetylglucosamine (GlcNAc)-binding site in F17G [15]. Although none of the residues in the UcaD/UclD pocket are conserved in F17G (Fig 4B and 4C) several of the residues could potentially form hydrogen bonds with a monosaccharide. The pocket also contains a tryptophan residue, which is known to mediate CH- π stacking interactions in the interaction site of other carbohydrate-binding proteins [23].

Next, we employed a glycan array to identify potential glycans that interact with UclD^{LD} and UcaD^{LD}. Of the 358 glycan structures present on the array, UclD^{LD} bound specifically to four glycans (Fig 5). Of these glycans, three were sialylated and three were α 1–3 fucosylated, with two structures, sialyl-Lewis X (Neu5Ac α 2-3Gal β 1-4(Fuc α 1-3)GlcNAc) and sialyllacto-N-fucopentose VI (Neu5Ac α 2-6Gal β 1-4GlcNAc β 1-3Gal β 1-4(Fuc α 1-3)Glc), having both sialylation and fucosylation. To further confirm and evaluate the binding affinities of UclD^{LD} with these glycans, surface plasmon resonance (SPR) analyses were undertaken (Fig 5). Overall, SPR analyses revealed the strongest binding interaction occurred between UclD^{LD} and sialyllacto-N-fucopentose VI ($K_D = 11.72 \text{ nM} \pm 3.6$). By contrast, UcaD^{LD} bound specifically to ten glycans from the glycan array, all but one of which, surprisingly, were different to the glycans bound by UclD^{LD} (Fig 5). SPR analysis revealed that UcaD had the highest binding affinity to Gal β 1-3GalNAc β 1-3Gal-sp4, Neu5Ac α 2-3(6-O-Su)Gal β 1-4(Fuc α 1-3)GlcNAc β -sp3, Gal β 1-4GlcNAc β 1-6(Gal β 1-4GlcNAc β 1-3)GalNAc α -sp3 and Neu5Ac α 2-8Neu5Ac α 2-8Neu5Ac α 2-3Gal β 1-4Glc-sp2. Both UclD and UcaD recognised lacto-N-fucopentose VI (Gal β 1-4GlcNAc β 1-3Gal β 1-4(Fuc α 1-3)Glc) at low micromolar affinities (Fig 5). Taken together, these data demonstrate that despite their high amino acid sequence identity, UclD and UcaD bind to different glycan oligosaccharides.

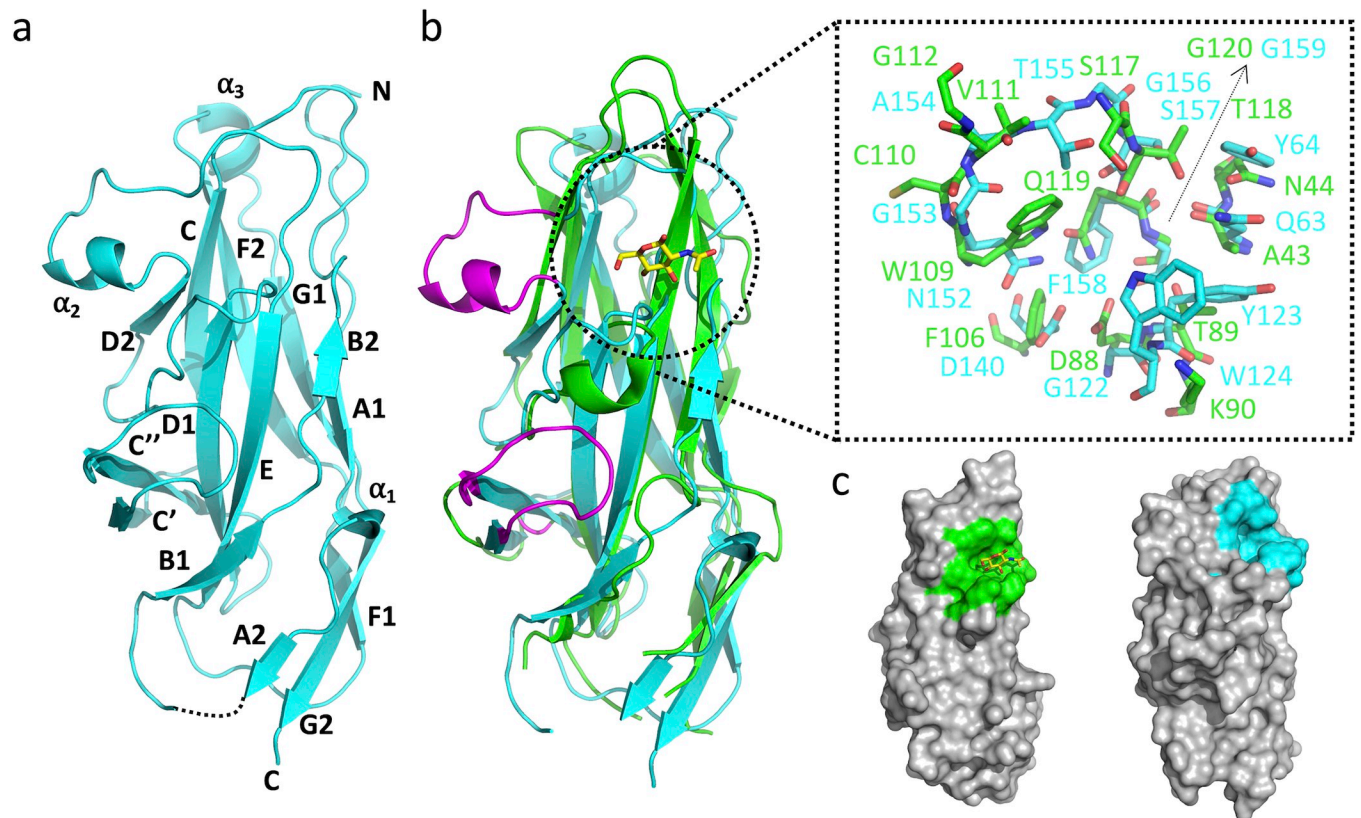


Fig 4. Crystal structure of UcaD^{LD}. a) Cartoon representation of UcaD^{LD}, with secondary-structure elements labelled A-G in accordance with previous descriptions [21]. b) Superposition of the UcaD^{LD} (cyan) and F17G (green, PDB: 1O9W) structures. The GlcNAc molecule bound to F17G is highlighted in stick representation. The insert shows the residues of the F17G GlcNAc-binding pocket, and the corresponding residues in UcaD^{LD}. The two sequence insertions observed in UcaD^{LD} and UclD^{LD}, compared to F17G, are highlighted in purple. c) Surface representation of F17G (left) and UcaD^{LD} (right), with the GlcNAc-binding pocket in F17G and corresponding pocket in UcaD^{LD} highlighted in green and cyan, respectively. The pockets were identified by comparing superimposed structures of F17G:GlcNAc, UcaD^{LD} and UclD^{LD}. The structures were superimposed using PyMOL.

<https://doi.org/10.1371/journal.ppat.1010582.g004>

Crystal structures of UcaD^{LD} bound to monosaccharides reveal an anchor point for glycan receptor binding via interaction with fucose and glucose residues

The putative carbohydrate binding pocket in the UclD^{LD} and UcaD^{LD} structures is solvent exposed in the UcaD^{LD} crystals (Fig O [a] in S1 Text), and adopts a near-identical conformation in three different UcaD^{LD} ligand-free crystal forms (PDB: 6H1X, 6H2L [21] and 7MZO [this study]; Fig O [b] in S1 Text), suggesting that the pocket is rigid and unlikely to significantly change conformation upon ligand binding. To determine if the binding pocket corresponds to a carbohydrate receptor binding site, we soaked pre-formed UcaD^{LD} crystals with the monosaccharides fucose (Fuc), glucose (Glc), galactose (Gal), N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc), and N-acetylneuraminic acid (Neu5Ac). Clear monosaccharide electron density was observed for the UcaD^{LD} crystals soaked with either Fuc, Glc, or Gal but not with GalNAc, GlcNAc or Neu5Ac. In all three cases, the monosaccharides were observed in the same binding pocket (Fig 6 and Table B in S1 Text).

The formation of the Fuc, Glc and Gal complexes results in a similar buried surface area of the monosaccharides (77.7–82.9%; 232.6–253.2 Å²), and superposition of the liganded and unliganded structures reveal that neither of the three sugars affect the conformation of the


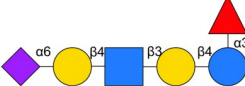
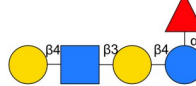
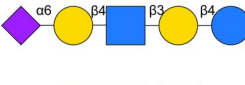
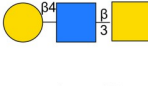
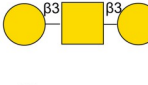
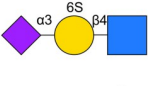
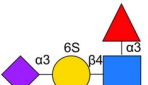
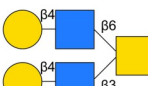
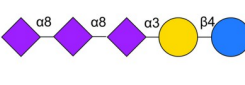
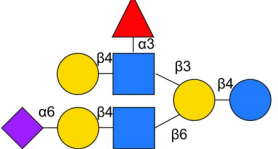

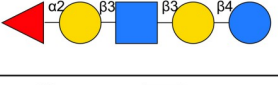
Name	Graphical structure	UclD	UcaD
SialylLewis X		2.52 $\mu\text{M} \pm 0.14$	No binding
Sialyllacto-N-fucopentose VI		11.7 nM ± 3.6	No binding
Lacto-N-fucopentose VI		1.48 $\mu\text{M} \pm 0.24$	6.47 $\mu\text{M} \pm 1.9$
LS-Tetrasaccharide c		1.97 $\mu\text{M} \pm 0.16$	No binding
Lactosamine-beta1-3Tn antigen		No binding	999.9 nM ± 269
TF-antigen-beta1-3Gal		No binding	39.19 nM ± 9.9
6-O-Sulfated-alpha2-3-Sialyl-lactosamine		No binding	135.2 nM ± 23
6-O-Sulfated-Sialyl-Lewis X		No binding	40.6 nM ± 7.2
Biantennary lactosamine-Tn-antigen		No binding	41.3 nM ± 7.3
Trisialo-ganglioside 3 (GT3)		No binding	52.3 nM ± 21
Monosialyl, monofucosyl-lacto-N-neo hexaose		No binding	549.6 nM ± 25
beta1-4galactosyl-galactose		No binding	125.7 nM ± 7.4
Lacto-N-fucopentose I		No binding	154.1 nM ± 9.9

Fig 5. Glycan interactions and binding affinity with UclD and/or UcaD. Shown are the glycans, together with their graphical structure and binding affinity to UclD^{LD} and/or UcaD^{LD} determined by SPR.

<https://doi.org/10.1371/journal.ppat.1010582.g005>

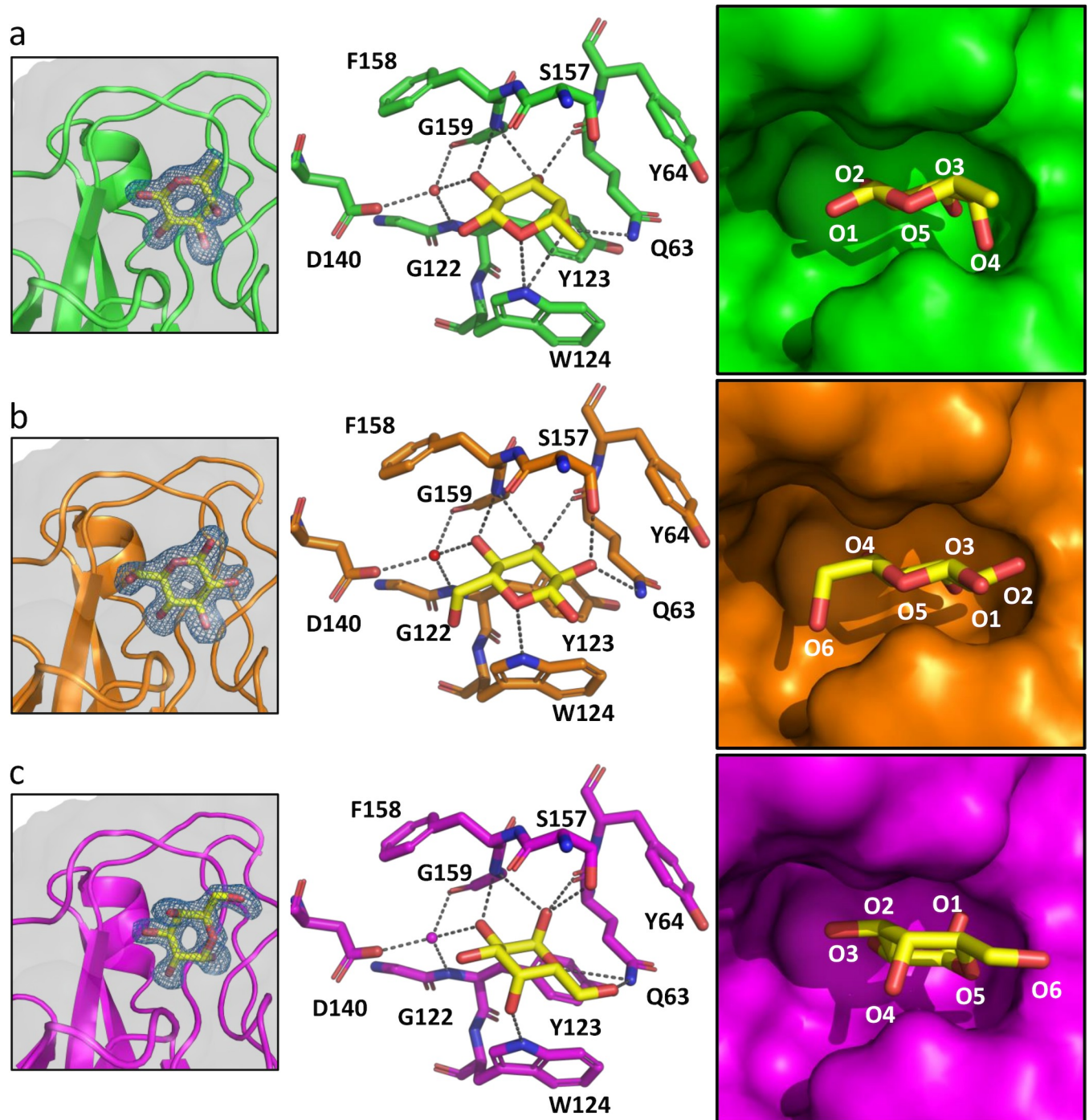


Fig 6. UcaD^{LD}-monosaccharide complexes. Binding site details of the, a) UcaD^{LD}:fucose, b) UcaD^{LD}:glucose, and c) UcaD^{LD}:galactose interactions. Left panel: Omit electron density maps of the ligands ($mF_o - DF_c$, blue mesh), contoured at 3.0σ . Middle panel: Bound monosaccharides and the residues implicated in recognition of monosaccharides are displayed in stick representation. Interactions between the monosaccharides and residues of the binding pocket of $\leq 3.6 \text{ \AA}$ are shown as dashed lines. Water molecules are shown as red spheres. Right panel: Surface representation of the binding sites. In the refined structure of the UcaD^{LD}:Fuc complex, positive difference density is observed adjacent to C4 (Fig P in S1 Text) suggesting that a minor fraction of the bound Fuc molecules adopt an alternate conformation in the crystal with the C4' hydroxyl interacting with S157. This conformation was not modelled in the deposited structure.

<https://doi.org/10.1371/journal.ppat.1010582.g006>

protein. In all three complex structures, the carbohydrate ligand interacts directly with residues Q63, G159 and W124 of the binding pocket, but the binding mode of each ligand is different (Fig 6 and Table B in S1 Text). The Fuc molecule is bound in the β -anomer configuration (β Fuc), with the C1' hydroxyl group exposed to solvent. The Glc molecule is flipped 180° compared to Fuc and adopts the β -configuration, with the C1' and C6' hydroxyl groups exposed to solvent, suggesting that the binding pocket can support binding to glycans that contains a terminal Glc residue, or either a reducing end or an internal Glc residue with C1 and C6 involved in glycosidic bonds. By contrast, the Gal C1' hydroxyl is buried; therefore, this Gal binding mode is not compatible with a glycan, because C1 cannot partake in a glycosidic bond. Furthermore, we did not observe binding of β -methyl-galactose in the UcaD^{LD} crystals, suggesting that the binding pocket cannot accommodate the β -anomer of galactose, which is the predominant configuration found in glycans.

Analysis of the UcaD residues involved in Fuc, Glc and Gal binding revealed that they are conserved in UclD. We also compared UclD, UcaD:ligand-free and UcaD:monosaccharide-bound structures, which demonstrated that the residues involved monosaccharide binding have a near identical conformation in the ligand-free structures of both UcaD and UclD (Fig O [c] in S1 Text). Thus, it is very likely that both UcaD and UclD bind in the same manner to the same set of monosaccharides.

Competition SPR and molecular docking of lacto-N-fucopentose VI to UcaD^{LD} reveals involvement of the DE loop in receptor binding

Our original aim was to define the precise molecular interaction between the interacting glycans and UclD^{LD}/UcaD^{LD}; however, apart from the receptor lacto-N-fucopentose VI (Gal β 1-4GlcNAc β 1-3Gal β 1-4(Fuc α 1-3)Glc), which is recognized by both adhesins at low-micromolar affinity, we were unable to source any of the interacting glycans in sufficient amounts for further structural analysis. We were unable to observe lacto-N-fucopentose VI in crystals with either UclD^{LD} and UcaD^{LD}, possibly due to steric hindrance by crystal contacts. As an alternative, we sought to determine if the monosaccharide found in the crystal structure bound to the same site on the protein as the glycans discovered using glycan array and SPR, by performing an SPR competition experiment using immobilised UcaD together with galactose and lacto-N-fucopentose VI (Fig 7). These studies revealed that pre-bound galactose ($K_D = 162.3\mu\text{M} \pm 12.8$) could inhibit lacto-N-fucopentose VI ($K_D = 1.48\mu\text{M} \pm 0.24$) binding, confirming that both ligands are engaged by a common lectin site, albeit at different affinities.

Next, we combined docking with molecular dynamics (MD) simulations to obtain further insight into the structural basis of glycan-protein interactions. The lacto-N-fucopentose VI receptor was selected for this analysis, since it is recognised by both UcaD and UclD at low-micromolar affinity. We used the UcaD^{LD}: β Fuc structure to guide the molecular docking of lacto-N-fucopentose VI, which was further optimized by a 40 ns MD simulation. During this MD simulation, the glycan remained bound to UcaD^{LD}, without dissociation for the entire duration of the simulation (Fig 8A). The α Fuc moiety of the receptor remained anchored in the monosaccharide-binding pocket (Fig 8A), but underwent a $\sim 90^\circ$ counter clockwise rotation compared to the crystal structure (Fig Q in S1 Text). This is not surprising, as the C1' hydroxyl group of the α Fuc molecule, based on the UcaD^{LD}: β Fuc structure, faces the protein surface, and therefore the α -1,3 linked glucose residue of the receptor cannot be accommodated by the lectin without a change in the fucose binding mode (Fig Q in S1 Text).

The DE loop, which has a large insertion in UclD^{LD} and UcaD^{LD} relative to F17G interacts with the terminal Gal and GlcNAc residues of lacto-N-fucopentose VI in the MD derived structure. This loop contains two variable residues; A143 and A150 in UcaD^{LD} correspond to a

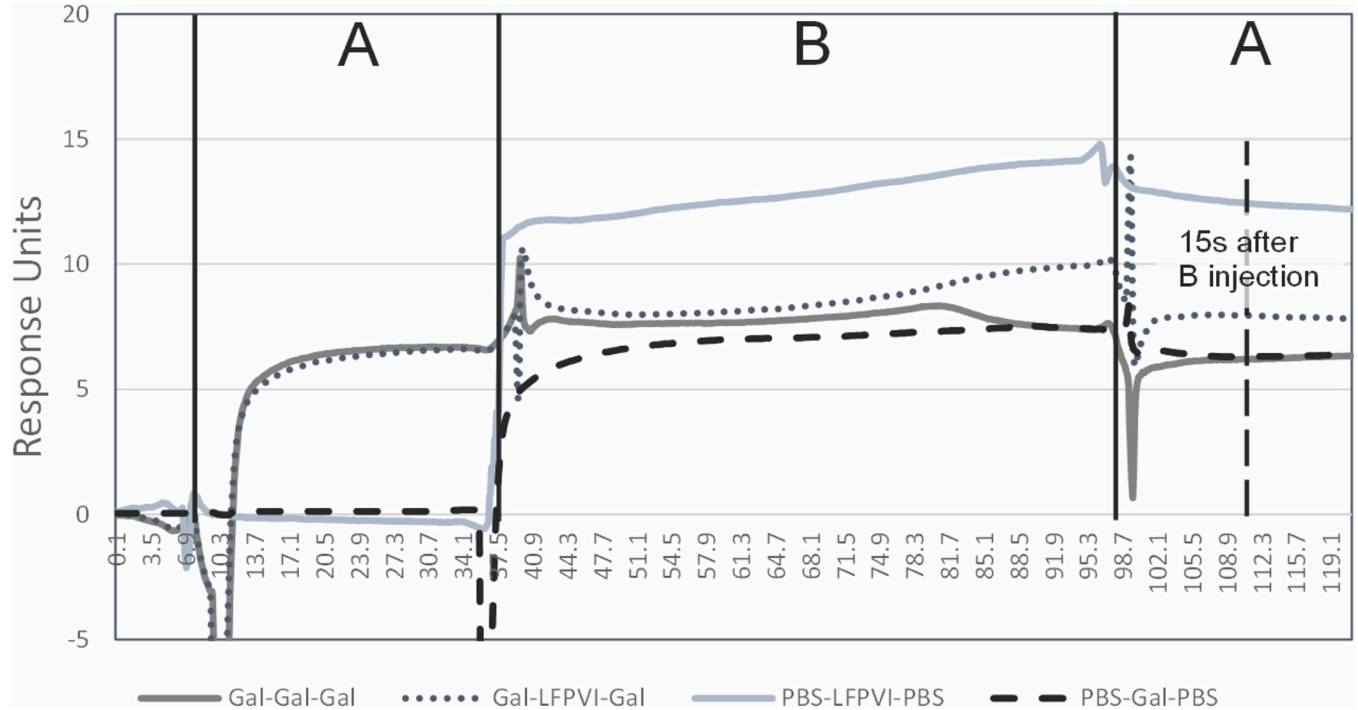


Fig 7. Competition between galactose and lacto-N-fucopentose VI binding to UclD. Injections are indicated as injection-A (PBS or galactose at 1 mM for 30 seconds) and injection-B (galactose at 1 mM or lacto-N-fucopentose VI at 50 μ M for 60 seconds). Final responses units (RU) of injections are recorded (indicated by dashed vertical line and label), with the response of galactose equalling $6.3RU \pm 0.15$ at 15 seconds after injection-B and lacto-N-fucopentose VI equalling $12.4 RU \pm 0.41$ at 15 seconds after injection-B. The response 15 seconds after injection-B when galactose and lacto-N-fucopentose VI were competed was $7.9 RU \pm 0.3$. As this value sits between 6.3 and 12.4 RU, it indicates direct competition for the same site with lacto-N-fucopentose VI replacing some of the bound galactose during the 60 second injection-B.

<https://doi.org/10.1371/journal.ppat.1010582.g007>

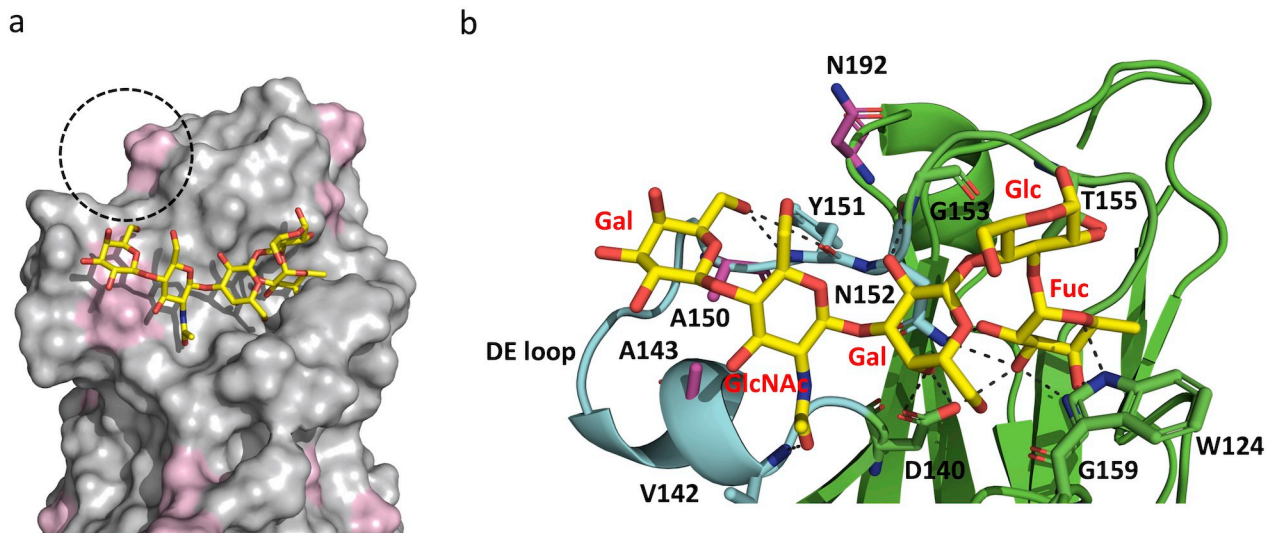


Fig 8. MD-derived structure of lacto-N-fucopentose VI bound to UcaD^{LD}. A representative minimum-energy conformation of the glycan (yellow) shown in stick representation. a) Surface representation of UcaD^{LD}. UcaD^{LD}/UclD^{LD} variable residues are highlighted in purple. The dashed circle highlights a potential binding site for the sialic acid residue of sialyllacto-N-fucopentose VI, which binds to UclD^{LD} with high affinity. b) Cartoon representation of UcaD^{LD} with the DE loop highlighted in cyan. Lacto-N-fucopentose VI-interacting residues and UcaD/UclD variable residues (purple) are shown in stick representation. In this model, the glycan is stabilised by polar interactions with the side-chains of D140 (internal Gal C4' hydroxyl) and T155 (Glc C2' hydroxyl), and the backbone of Y151 (terminal Gal C6' hydroxyl), V142 (GlcNAc amine group) and G153 (internal Gal C2' hydroxyl).

<https://doi.org/10.1371/journal.ppat.1010582.g008>

glycine and threonine, respectively, in UclD^{LD} (Fig 8A and 8B). The differences in surface properties caused by these two amino acid substitutions could impact glycan specificity and affinity. The addition of a sialic acid residue to the lacto-N-fucopentose VI receptor via α 2.6 glycosidic bond to the terminal Gal increases the binding affinity for UclD^{LD} by two orders of magnitude, while no binding is detected for UcaD^{LD} (Fig 5). Based on the docked UcaD^{LD}: lacto-N-fucopentose VI complex, the sialic acid residue could reside in close proximity to another surface region with variable UclD^{LD}/UcaD^{LD} regions (Fig 8A). In particular, the swap of asparagine 192 in UcaD^{LD} to a positively charged lysine residue in UclD^{LD} could be the reason for the increased binding affinity for sialyllacto-N-fucopentose VI via its negatively charged sialic acid.

In summary, our structural studies reveal a monosaccharide-binding pocket with broad specificity, recognising sugars such as Glc, Fuc and Gal, but not amino sugars such as sialic acid, GlcNAc and GalNAc, which may be due to the inability of the pocket to accommodate the more bulky N-acetyl amine group. Only fucose binding is consistent with the identified UcaD and UclD glycan receptors, but glycan anchoring to these lectins via fucose is likely to require a different binding mode compared to the UcaD:Fuc crystal structure. Our SPR competition experiments and molecular docking analysis suggests that the extended DE loop, which is a unique feature of the UcaD and UclD adhesins, plays an important role in engaging with fucose-containing glycan receptors.

Discussion

Adherence is a pivotal initial step in bacterial colonisation and disease pathogenesis. Fimbriae of the CU secretion pathway represent one of the primary factors for mediating colonisation of specific host tissues. Here, we describe a mechanism of ExPEC adaptation to intestinal colonisation via enhanced expression of Ucl fimbriae, and complement this with a detailed description of the Ucl fimbriae glycan receptor repertoire as well as the identification of a broad-specificity glycan binding pocket in the UclD fimbrial adhesin that could accommodate these interactions.

The genes encoding Ucl fimbriae were found within a limited range of *E. coli* STs, and predominantly within STs from the B2 phylogenetic group. This variable prevalence extends previous studies [9, 10, 12], and complete sequencing of the F11 genome defined the precise location of the *ucl* genes within *GI-leuX*. Bioinformatic analysis also demonstrated allelic variation of the major subunit *uclA* gene, providing evidence for immune pressure driving antigenic variation, as observed for other fimbriae such as type 1 and P fimbriae [24, 25]. Comparative analysis between F11 and several other UPEC strains, including UTI89, revealed a striking difference in the level of expression of the UclA major subunit protein, and this correlated with a SNP in the regulatory region that impacts promoter activity. SNPs that alter host colonisation have mostly been described within coding regions, an example being minor pathoadaptive amino acid variations in the FimH adhesin of type 1 fimbriae that enhance UPEC colonisation of the bladder in experimental mice [26, 27]. SNPs within non-coding regulatory regions, albeit less well-characterised, can also enhance host colonisation by altering expression of key virulence determinants [28]. Here, we show the T/G⁽⁻⁷⁸⁾ nucleotide in the *ucl* regulatory region impacts transcriptional activation by the global regulator OxyR, and that T⁽⁻⁷⁸⁾ SNP in strain F11 drives increased colonisation of the mouse intestine. OxyR was shown to bind to both the *Pucl*^{-78G} and *Pucl*^{-78T} sequences, with higher affinity to *Pucl*^{-78G}, consistent with its closer match to the consensus OxyR binding site, and suggesting the regulation of Ucl fimbriae is more complex and likely to involve other regulatory cofactors. We also show that SNPs in the OxyR binding site occur more broadly in ST127, and that at least one of these

additional SNPs also causes increased UclA expression. Thus, our data provides strong evidence that mutations within this regulatory region, together with transcriptional control by OxyR and other factors yet to be identified, represent a mechanism that enables *E. coli* to tune Ucl fimbriae expression.

OxyR is a global transcriptional regulator that controls cellular responses to oxidative stress [29]. Our data provide the first description of OxyR regulated fimbriae expression in *E. coli*, an important finding given the enormous diversity of CU fimbriae. OxyR has been shown to activate the expression of fimbriae in other organisms [30, 31], and thus this mode of regulation may define a new paradigm linking fimbrial expression to oxidative stress and gut colonisation. Although oxygen availability decreases through the intestine and the large intestine is essentially anaerobic, higher levels of oxygen are found close to the gut mucosa [32]. Therefore, in line with a previous study that showed purified UcaD^{LD} binds to colonic epithelial cells in tissue sections of the mouse gut [12], it is likely that ExPEC may be exposed to oxidative stress at the gut mucosa, causing increased expression of Ucl fimbriae and thus driving enhanced colonisation. Another mechanism of OxyR activation in the gut occurs through bile, which activates stress responses that may promote interaction of *E. coli* with cells of the colonic epithelium [33]. Bile can also act as a mutagen, and in *Salmonella* bile exposure increases the frequency of G-C to A-T mutations, which results in enhanced expression of genes regulated by OxyR and SoxRS [34], posing an intriguing link to our observations.

The phylogenetic relatedness and role in pathogenesis of UPEC Ucl fimbriae and *P. mirabilis* Uca fimbriae led us to study the properties of both fimbrial adhesins. We evaluated the glycan-binding specificities of UclD^{LD} and UcaD^{LD} against glycan arrays containing 358 structures that represent host-cell surface glycoconjugates. Our analysis identified four glycans that interact with UclD and nine glycans that interact with UcaD, with all interactions confirmed and quantified by SPR. Overall, both UclD and UcaD displayed different oligosaccharide-binding specificities, with only one glycan, lacto-N-fucopentose VI, bound by both adhesins. The N-acetyllactosamine (Gal β 1-4GlcNAc) motif was present in several glycans, suggesting that this oligosaccharide interacts with both UclD and UcaD. Previous studies have shown that purified Uca fimbriae bind to asialo-GM1, asialo-GM2 and lactosyl ceramide glycolipids, and that these interactions are likely to occur via the UcaA major subunit rather than the UcaD adhesin [35]. Consistent with these reports, we did not identify interactions between UclD^{LD} and these glycans.

The glycans identified for UclD are all structurally similar, featuring a sub-terminal α 1-3 fucosylation and/or a terminal sialylation on a Gal β 1-4GlcNAc structure. The highest affinity glycan, sialyllacto-N-fucopentose VI (Neu5Ac α 2-6Gal β 1-4GlcNAc β 1-3Gal β 1-4(Fuc α 1-3)Glc), contains both the fucose and sialic acid motif with a terminal α 2-6Neu5Ac. Interestingly, the same structure lacking the fucose LS-tetrasaccharide c (Neu5Ac α 2-6Gal β 1-4GlcNAc β 1-3Gal β 1-4Glc) or without the sialic acid lacto-N-fucopentose VI (Gal β 1-4GlcNAc β 1-3Gal β 1-4(Fuc α 1-3)Glc) was also recognised, but both had 100-fold lower affinity compared to the complete structure (Fig 5). Sialyllacto-N-fucopentose VI was first discovered in human milk as a human milk oligosaccharide (HMO) [36]. HMOs play a key role in protecting infants from infections by mimicking host glycan cellular receptors for bacterial and viral gut pathogens, consistent with a role for sialyllacto-N-fucopentose VI as a receptor for Ucl in the human gut. Binding for UcaD showed little direct overlap with UclD, with the highest affinity binding identified for UcaD being to neutral glycans containing GalNAc on the glycan core and sulfated Gal β 1-4GlcNAc structures.

Our structural data identified a monosaccharide-binding pocket in the UclD and UcaD lectin-binding domains that can act as an anchor point for interaction with glycan receptors, consistent with previous predictions [12, 21]. Although the binding pocket is localised in the same

structural region as the GlcNAc-binding pocket in the F17G-like adhesins [15], the residues involved in monosaccharide binding are not conserved. Our crystal-soaking studies with various monosaccharides revealed that the pocket can accommodate multiple monosaccharide binding modes and interacts with sugars such as glucose, fucose and galactose, but not amino sugars such as sialic acid, GlcNAc, and N-acetylgalactosamine, which is probably due to the inability of the pocket to accommodate the more bulky N-acetyl amine group. The broad specificity is consistent with the ability of these lectins to recognise multiple glycan receptors. Our molecular docking and MD analysis suggests that the fucose residue of fucose-containing UcaD and UclD receptors can be anchored to this pocket, but has to adopt a different binding mode compared to the crystal structure in order for the additional components of these receptors to be accommodated by the lectins (Figs 8 and Q in S1 Text). Finally, our SPR competition experiments and molecular-docking analysis also suggests that the extended DE loop, which is a unique feature of the UcaD and UclD adhesins, plays an important role in engaging with fucose-containing glycan receptors.

In summary, we report a detailed molecular, structural and functional characterisation of *E. coli* Ucl fimbriae. We demonstrate how minor mutations in the promoter regulatory region enable tuning of Ucl fimbrial expression mediated via the global OxyR regulator, and reveal the impact of such changes on intestinal colonisation in a mouse model. Long-term persistence in the human gut despite ongoing antibiotic treatment can support the survival of ExPEC reservoirs that seed recurrent UTI [37]. Thus, blocking bacterial adherence, as shown in preclinical models for UPEC type 1 and Fml fimbriae [12, 38], represents an attractive non-antibiotic approach to address the global threat of rapidly increasing antibiotic resistance. The identification of glycan receptor targets for the UclD adhesin provides a framework to understand Ucl fimbriae tissue tropism and to develop new therapeutics that could be used as anti-adhesion drugs.

Materials and methods

Ethics statement

Mice were handled in accordance with protocols approved by the Institutional Animal Care and Use Committee at the University of Utah (protocol number 19–01001), according to U.S. federal guidelines indicated by the Office of Laboratory Animal Welfare and described in the Guide for the Care and Use of Laboratory Animals, 8th ed.

Bacterial strains, mutants and culture conditions

Bacterial strains and plasmids used in this study are listed in Table C in S1 Text. Gene disruption mutants were generated using λ -Red recombinase-mediated homologous recombination as previously described [39, 40]; single-nucleotide switching of F11 and UTI to F11-*PucI*^{T-78G} and UTI89-*PucI*^{G-78T} was performed using pORTMAGE vectors as described previously [18, 41]. Primers used for the generation of mutants are listed in Table D in S1 Text. Bacteria were routinely grown at 37°C on solid or in liquid LB media supplemented with appropriate antibiotics ampicillin (100 µg/ml), kanamycin (25 µg/ml) or chloramphenicol (30 µg/ml).

Bioinformatic and molecular techniques

Bioinformatic analyses were performed as described in the Supplementary Methods in S1 Text. The F11 genome was completely sequenced using a combination of long-read Pacific Biosciences Single Molecule Real-Time (SMRT) and short-read Illumina sequencing as previously described [42,43]. Methods for DNA manipulation, rapid amplification of cDNA ends

(5' RACE), mutant construction, protein preparation, immunoblotting, whole-cell ELISA and β -galactosidase assays were performed as described in the Supplementary Methods in [S1 Text](#).

Microscopy

Bacterial cells for electron microscopy were fixed with 2.5 % glutaraldehyde. Glow-discharged carbon-coated Formvar copper grids were placed on drops of bacterial suspension for 1 min and then washed on drops of water (3 x 1 min). Grids were negatively stained with 0.5% uranyl acetate and cells were examined under a JEOL 1010 transmission electron microscope operated at 80 kV. Micrographs were captured using an analysis Megaview III digital camera. For immuno-gold electron microscopy, fixed bacterial cells (10 μ L) were spotted onto glow-discharged carbon-coated Formvar copper grids and probed with primary antibodies (1:100 in 1% BSA/PBS) and secondary species-specific gold-labelled conjugates (5- or 10-nm size particles) and then negatively stained with 0.5% uranyl acetate.

Electrophoretic mobility shift assays

Plasmid pOxyR was modified by PCR to introduce an in-frame sequence encoding glycine-glycine-6xhistidine. Recombinant OxyR-6xHis was purified using nickel affinity chromatography; gel shift assays were performed as previously described [44]. Briefly, purified OxyR was incubated with DNA in a 25 μ L reaction containing binding buffer (20 mM Tris (pH 7.5), 50 mM KCl, 1 mM EDTA, 10% glycerol v/v) at room temperature for 15 minutes. Cy3-labelled DNA was mixed with purified protein before addition of unlabelled DNA, where appropriate. The reactions were run in 4.5% native acrylamide (Bio-Rad, catalogue no. 1610146) in 0.5 \times Tris-borate-EDTA (TBE) at 4°C, 90 V for one hour. Cy3-labelled DNA was visualized using an Amersham imager 800 (Cytiva, catalogue no. 29399481).

Mouse gut colonization assays

Single and mixed competitive gut colonization assays were carried out as previously described [12, 45], with detailed methods provided in the Supplementary Methods in [S1 Text](#). For mixed infection assays, seven- to eight-week-old female Specific Pathogen Free C3H/HeN mice (Charles River Labs) were initially treated with streptomycin, and then inoculated with 50 μ L PBS containing a total of $\sim 5 \times 10^7$ CFU, comprised of a 1:1 mix of F11::kan and F11::cm (control) or F11::kan and F11-PucI^{T-78G} (cm resistant). At the indicated time points post-inoculation, individual mice were briefly (3 to 10 min) placed into unused takeout boxes for weighing and feces collection. Competitive indices (CI) were calculated as the ratio of the cm resistant over kan resistant bacteria recovered from the feces divided by the ratio of the same strains within the inoculum.

UclD^{LD} and UcaD^{LD} expression and structural analysis

The coding sequences for the *uclD* and *ucaD* lectin domains were amplified from *E. coli* strain F11 and *P. mirabilis* PM54 clinical isolate [46], respectively. The amplicons were cloned into pET22b, which encodes a C-terminal His-tag. Epoch Life Science made and confirmed by sequencing the pET22b::*uclD* and pET22b::*ucaD* constructs. The expression and purification of UclD^{LD} and UcaD^{LD} was performed as described in the Supplementary Methods in [S1 Text](#). Crystals were produced by the hanging drop method. Structural analyses were carried out using DALI [22], and PyMOL (<http://www.pymol.org>). Figures were generated using PyMOL. The UclD^{LD} structure was refined to final $R_{\text{work}}/R_{\text{free}}$ values of 24.8%/30.6%; the UcaD^{LD} structure was refined to final $R_{\text{work}}/R_{\text{free}}$ values of 16.9%/19.2%, respectively (Table A in [S1](#)

Text). For UcaD^{LD}:monosaccharide complexes, coordinates and structure factors have been deposited in the PDB with IDs 7MZQ (Fuc complex), 7MZR (Glc complex), and 7MZS (Gal complex). Detailed methods for crystallisation and structure determination of UclD^{LD}, UcaD^{LD} and monosaccharide complexes, as well as methods for molecular docking and molecular dynamics simulations, are described in the Supplementary Methods in [S1 Text](#).

Glycan array analysis

Glycan array analysis was carried out using methods previously described [47,48], with specific details outlined in the Supplementary Methods in [S1 Text](#). The full list of glycans analysed and the array analysis data is shown in Table E in [S1 Text](#). Slides were scanned and analysed as outlined in the MIRAGE compliance table (Table F in [S1 Text](#)). SPR analysis was carried out using a Biacore T200 system (Cytivia) as previously described [47] with minor modifications outlined in the Supplementary Methods in [S1 Text](#).

Competitive surface plasmon resonance

Competition assays (ABA-injection method according to the manufacturer's instructions; Cytiva S200) were performed between the galactose monosaccharide identified by crystallography and the best binding structure identified through glycan array and SPR analysis, lacto-N-fucopentose VI. UclD was immobilized as described above and competition analysis was performed using the ABA injection option within the kinetic/affinity methods builder. In this method, an injection molecule A at the start and end (30 seconds each injection) of the cycle and injection of a second molecule B (60 second injection) in between the two A injections. For this analysis PBS and galactose were used as 'A' injections, while PBS, galactose and lacto-N-fucopentose VI were used as 'B' injections. Galactose was used at 1 mM to ensure saturation, while lacto-N-fucopentose VI was used at 50 μ M (~10 times the measured K_D). Final responses units (RU) of injections are recorded with the response of galactose equalling ~6.3 RU 15 seconds after injection-B and lacto-N-fucopentose VI equalling ~12.4 RU 15 seconds after injection-B. If galactose and lacto-N-fucopentose VI compete for the same site then the RU 15 seconds after the injection when galactose and lacto-N-fucopentose VI is competed should be between 6.3–12.4 RU. If the response is higher than 12.4 RU then the two sugar molecules bind to different sites on the protein.

Accession codes

The sequences for the F11 chromosome and pF11A have been deposited in the NCBI GenBank database under accession numbers CP076123 to CP076124, respectively. Raw reads for the F11 genome have been deposited to the Sequence Read Archive with accession codes SRR14581595 and SRR14581594 for Illumina and Pacific Biosciences SMRT reads, respectively. Coordinates and structure-factor files have been deposited in the Protein Data Bank (PDB), with accession codes 7MZO (UcaD), 7MZP (UclD), 7MZQ (UcaD:Fuc complex), 7MZR (UcaD:Glc complex) and 7MZS (UcaD:Gal complex).

Supporting information

S1 Text. Supplementary information comprising Supplementary Methods (DNA manipulation and genetic techniques, Bioinformatic analysis, Genetic mutagenesis of bacteria, Protein preparation, immunoblotting and whole-cell ELISA, β -galactosidase assay, 5' RACE, Mouse gut colonization assays, Cloning, expression and purification of UclDLD and UcaDLD, Glycan array analysis, SPR Analysis, Crystallization and crystal structure

determination, and Molecular docking and molecular dynamics simulations); Supplementary Figures (Figs A-Q); Supplementary Tables (Tables A-F); and Supplementary Section references. Fig A. The Ucl fimbrial operon is most frequently found in phylogroup B2 strains. Percentage of strains encoding *uclABCD* for each sequence type in the 83ST database. The bars are split to show the presence of specific *uclA* alleles; the least common *uclA* variants are summarised by the “other” designation (other: 12 variants observed; 33/404, ~8%); see Fig B in S1 Text. The *ucl* genes were most frequently found in ExPEC strains from ST12, ST73, ST127, ST131 and ST141 in the pathogenic B2 phylogroup (360/900; 40%), compared to strains from STs in phylogroups D (15/600; 2.5%) and F (3/500; 0.6%), and rarely found in strains from STs in other *E. coli* phylogroups. Fig B. Maximum likelihood phylogeny of *uclA* variants found in the 83ST database, with the number of each *uclA* variant indicated in the bar graph. A total of 15 *uclA* allelic variants that differed by up to 27% at the nucleotide level were identified. The most common allelic variant was *uclA*-10 (212/404; 52.5%), followed by *uclA*-5 (88/404, 21.8%) and *uclA*-13 (71/363; 17.6%); other allelic variants were infrequent. Fig C. Pair-wise sequence comparison of GI-*leuX* region of UTI89 (top), F11 (middle) and MG1655 (bottom). The integase gene adjacent to tRNA-*leuX* is coloured blue, *uclABCD* are coloured red. The grey regions connecting the genomes represent blastn output with the percent of conservation shown in the scale bar. Fig D. Analysis of the *ucl* promoter. A) Left, schematic diagram of the *ucl* promoter region from F11 and UTI89 cloned in the reporter plasmid pQF50. Indicated are the TSS, -10 and -35 promoter elements, and OxyR binding site. Right, β -galactosidase activity (measured in Miller units) for each *Pucl-lacZ* fusion construct in UTI89*lacZ*. Plasmid pQF50-*Pucl*^{F11}-*lacZ* possessed a higher β -galactosidase activity as compared to pQF50-*Pucl*^{UTI89}-*lacZ* ($p < 0.0001$; one-way ANOVA with Sidak’s multiple comparisons test). B) Promoter region of *ucl* operon from F11. The transcription start site is indicated as +1, with the predicted -10 and -35 core promoter elements indicated accordingly. Bolded T with an arrow indicates the single nucleotide that differed in F11, where a G is present at this position in UTI89, S77EC and HVM1299. The region from F11 and UTI89 cloned into the pQF50 *lacZ* reporter plasmid is indicated by arrows denoting the 5’ ends of primers 6571–6572. Fig E. 5’ RACE of *uclA* to map the transcription start site. Top: sequence, indicating the ATG start codon of the *uclA* gene and the +1 transcription start site. Bottom: sequence chromatogram. Fig F. a) Whole-cell lysate western-blot analysis of F11^{T-78G} and UTI89^{G-78T}. Higher expression of UclA was observed in UTI89^{G-78T} mutants and lower in abundance in F11^{T-78G}, compared to that of their wild-type strains. b) Qualitative analysis of 100 cells assessed for Ucl fimbriation using α -UclA immuno-gold labelling. c) Representative UclA immunogold-labelled TEM images for F11, F11 Δ *uclA* and F11-*Pucl*^{T-78G}. Fig G. Anti UclA whole-cell ELISA. a) Whole-cell ELISA demonstrating expression of Ucl fimbriae on wild-type F11 (wt), F11 Δ *ucl* (*ucl*) and F11-*Pucl*^{T-78G} (T-78G), as well as wild-type UTI89 (wt), UTI89 Δ *ucl* (*ucl*) and UTI89-*Pucl*^{G-78T}. Ucl fimbriae were detected using UclA-specific polyclonal antibody. b) Control whole cell ELISA of wild-type F11 (wt), F11 Δ *ucl* (*ucl*) and F11-*Pucl*^{T-78G} (T-78G), as well as wild-type UTI89 (wt), UTI89 Δ *ucl* (*ucl*) and UTI89-*Pucl*^{G-78T} employing a general *E. coli* antibody (Life Research B65001R). The black dots show individual measurements for four technical replicates from three biological replicates ($n = 12$); the grey bar indicates the mean. Statistical analyses were performed by one-way ANOVA with Sidak’s multiple comparisons test. Fig H. Binding of OxyR to the *ucl* promoter region (*Pucl*). a, Coomassie-stained SDS-PAGE of OxyR-6xHis. Lane M: PageRuler Prestained Protein Ladder (Life Technologies, catalogue no. 26616), Lane 2: Nickel-affinity purified OxyR-6xHis. b, Schematic diagram of the *uclABCD* operon, indicating the transcription start site (TSS), -10 and -35 promoter region, and OxyR binding sequence containing the T⁽⁻⁷⁸⁾ in F11 and the G⁽⁻⁷⁸⁾ in UTI89. Also indicated are the 261 bp *Pucl* PCR fragment and the 240 bp *uclC* PCR fragment (negative

control) used in the gel shift assay. c, Electrophoretic mobility shift assay of the Cy3-*Pucl*^{-78T} (top) and Cy3-*Pucl*^{-78G} (bottom) fragments with OxyR and increasing concentrations of unlabelled competitor (*Pucl*-UL) DNA. Fig I. Conservation of the *Pucl* OxyR binding site in ST127. A total of 845 genomes from ST127 strains on Enterobase were downloaded, 698 of which were positive for the Ucl fimbriae genes. The OxyR binding site was extracted from each *Pucl* sequence and aligned to generate the DNA logo shown at the top of the figure. Twenty-seven unique OxyR binding sites were identified, with nucleotide sequence changes shown below the consensus sequence. The number of times each unique sequence was identified in the dataset is indicated. The OxyR binding site consensus sequence was found most frequently (n = 486), while the F11 T-78G SNP was also common (n = 96). In total, 31% of the *Pucl* OxyR binding sites contained at least one SNP compared to the consensus sequence, with the F11 T-78G SNP most prevalent (14%). Sequences containing the G-78T and T-76C are indicated. Fig J. UclA expression is increased in ECOR60. a) OxyR binding site sequences from UTI89 (consensus), F11 and ECOR60, highlighting the C-76T nucleotide sequence change in ECOR60. b) Whole-cell lysate western-blot analysis of ECOR60, ECOR63, F11 (WT) and F11-*Pucl*^{T-78G} employing a UclA-specific antibody. Higher expression of UclA was observed in ECOR60 and F11 (WT) compared to ECOR63 and F11-*Pucl*^{T-78G}. The promoter region of the *ucl* operon from ECOR63 and UTI89 is identical. Fig K. Ucl fimbriae do not impact colonisation of the mouse gut in single infection experiments. Mice were inoculated wild-type F11, F11-*Pucl*^{T-78G} and an F11Δ*ucl* mutant. Each group contained 10–11 mice infected and monitored during two independent experiments. Bacterial loads were assessed over a 2-week period. Fig L. Architecture of Ucl fimbriae demonstrated by co-immunogold labelled electron microscopy. a) Electron micrograph demonstrating immunogold labelled UclA major subunit (left; 5 nm gold particles) and UclD tip adhesin (right; 10 nm gold particles) of Ucl fimbriae. b) Cartoon model of Ucl fimbriae architecture, depicting the UclA major subunit repeating protein (green), UclD tip adhesin (orange), UclC usher (yellow) and UclB chaperone (red). Also labelled are the inner membrane (IM) and outer membrane (OM) of the cell. Fig M. Prevalence of the *uclABCD* genes in *Proteus* species. Genomes were assessed from the NCBI database. The analysis was performed using tblastn for UcaA, UcaB, UcaC and UcaD, with a positive result determined for blast hits with >70% identity and >80% coverage for all four proteins. Fig N. Amino acid sequence alignment of the UclD, UcaD and GafD adhesins. Identical amino acids are shaded in black; similar amino acids are shaded in grey. The consensus sequence is indicated. Fig O. A, Crystal packing analysis of UcaD^{LD} (PDB: 7MZO). The molecules are shown in cartoon representation. The asymmetric unit, which consists of one UcaD^{LD} molecule, is shown in cyan, while symmetry-related molecules are colored orange. The monosaccharide binding region of the asymmetric unit is circled and highlighted in magenta. B, Comparison of the monosaccharide binding site region in 3 different crystal forms of ligand-free UcaD^{LD}. C, Comparison of the monosaccharide binding site in the UcaD^{LD}:Fuc, UcaD^{LD} and UclD^{LD} crystal structures. Fig P. Electron density map of the fucose binding site in the UcaD^{LD}:Fuc complex structure. Composite (2Fo—Fc, blue, contoured at 2σ) and difference (Fo—Fc, green/red, contoured at 3.0σ) electron density map of the Fuc binding site after refinement. Positive difference density adjacent to C4 suggests that a minor fraction of the Fuc molecules in the crystal adopts an alternate conformation. Fig Q. UcaD^{LD}:Fuc interactions in the MD derived structure. Interactions between fucose (yellow) and residues of the binding pocket of ≤3.6 Å are shown as dashed lines. The binding model of the fucose molecule observed in the UcaD^{LD}:Fuc crystal structure is highlighted in green stick representation, for comparison. Table A. Crystallographic data. Table B. Polar interactions in the UcaD^{LD}: monosaccharide complexes. Table C. List of strains and plasmids used in this study. Table D. List of primers used in this study. Table E. Glycan array analysis of UclD and UcaD. Table F. Supplementary

glycan microarray document based on MIRAGE guidelines (DOI: [10.1093/glycob/cww118](https://doi.org/10.1093/glycob/cww118). (PDF)

S1 Data. F11 genomic analyses and results. This file contains information about the genome features, virulence factors, putative phage and methylome motif summary of F11. (XLSX)

Acknowledgments

We thank Xinying Jia and Mehdi Mobli from the Centre for Advanced Imaging at The University of Queensland for helpful discussions, Sohinee Sarkar for technical support and Rick Webb for expert microscopy assistance. We acknowledge the use of the University of Queensland Remote Operation Crystallization and X-ray (UQROCX) Facility at the Centre for Microscopy and Microanalysis and the support from staff, Gordon King and Karl Byriel. We also acknowledge use of the Australian Synchrotron MX beamlines and thank the staff for their support.

Author Contributions

Conceptualization: Steven J. Hancock, Alvin W. Lo, Thomas Ve, Christopher J. Day, Michael P. Jennings, Matthew A. Mulvey, Bostjan Kobe, Mark A. Schembri.

Data curation: Steven J. Hancock, Alvin W. Lo, Thomas Ve, Christopher J. Day.

Formal analysis: Steven J. Hancock, Alvin W. Lo, Thomas Ve, Christopher J. Day, Mark A. Schembri.

Funding acquisition: Michael P. Jennings, Matthew A. Mulvey, Bostjan Kobe, Mark A. Schembri.

Investigation: Steven J. Hancock, Alvin W. Lo, Thomas Ve, Christopher J. Day, Lendl Tan, Alejandra A. Mendez, Minh-Duy Phan, Nguyen Thi Khanh Nhu, Kate M. Peters, Amanda C. Richards, Brittany A. Fleming, Chyden Chang, Dalton H. Y. Ngu, Brian M. Forde, Thomas Haselhorst, Kelvin G. K. Goh.

Methodology: Steven J. Hancock, Alvin W. Lo, Thomas Ve, Christopher J. Day.

Project administration: Mark A. Schembri.

Resources: Michael P. Jennings, Matthew A. Mulvey, Bostjan Kobe.

Supervision: Scott A. Beatson, Michael P. Jennings, Matthew A. Mulvey, Bostjan Kobe, Mark A. Schembri.

Visualization: Thomas Ve, Christopher J. Day, Thomas Haselhorst, Matthew A. Mulvey.

Writing – original draft: Steven J. Hancock, Thomas Ve, Christopher J. Day, Thomas Haselhorst, Mark A. Schembri.

Writing – review & editing: Steven J. Hancock, Alvin W. Lo, Thomas Ve, Christopher J. Day, Minh-Duy Phan, Nguyen Thi Khanh Nhu, Scott A. Beatson, Michael P. Jennings, Matthew A. Mulvey, Bostjan Kobe, Mark A. Schembri.

References

1. Kaper JB, Nataro JP, Mobley HLT. Pathogenic *Escherichia coli*. *Nat Rev Microbiol*. 2004; 2(2):123–40. <https://doi.org/10.1038/nrmicro818> WOS:000220431700015. PMID: 15040260

2. Zowawi HM, Harris PN, Roberts MJ, Tambyah PA, Schembri MA, Pezzani MD, et al. The emerging threat of multidrug-resistant Gram-negative bacteria in urology. *Nature rev Urology*. 2015; 12(10):570–84. Epub 2015/09/04. <https://doi.org/10.1038/nrurol.2015.199> PMID: 26334085.
3. Manges AR, Geum HM, Guo A, Edens TJ, Fiske CD, Pitout JDD. Global extraintestinal pathogenic *Escherichia coli* (ExPEC) lineages. *Clinical microbiology rev*. 2019; 32(3). <https://doi.org/10.1128/CMR.00135-18> PMID: 31189557; PubMed Central PMCID: PMC6589867.
4. Denamur E, Clermont O, Bonacorsi S, Gordon D. The population genetics of pathogenic *Escherichia coli*. *Nat Rev Microbiol*. 2021; 19(1):37–54. <https://doi.org/10.1038/s41579-020-0416-x> PMID: 32826992.
5. Waksman G, Hultgren SJ. Structural biology of the chaperone-usher pathway of pilus biogenesis. *Nat Rev Microbiol*. 2009; 7(11):765–74. <https://doi.org/10.1038/nrmicro2220> PMID: 19820722; PubMed Central PMCID: PMC3790644.
6. Hospenthal MK, Costa TRD, Waksman G. A comprehensive guide to pilus biogenesis in Gram-negative bacteria. *Nat Rev Microbiol*. 2017; 15(6):365–79. <https://doi.org/10.1038/nrmicro.2017.40> PMID: 28496159.
7. Chen SL, Hung CS, Xu J, Reigstad CS, Magrini V, Sabo A, et al. Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103(15):5977–82. <https://doi.org/10.1073/pnas.0600938103> PMID: 16585510; PubMed Central PMCID: PMC1424661.
8. Klein RD, Hultgren SJ. Urinary tract infections: microbial pathogenesis, host-pathogen interactions and new treatment strategies. *Nat Rev Microbiol*. 2020; 18(4):211–26. Epub 2020/02/20. <https://doi.org/10.1038/s41579-020-0324-0> PMID: 32071440; PubMed Central PMCID: PMC7942789.
9. Wurpel DJ, Beatson SA, Totsika M, Petty NK, Schembri MA. Chaperone-usher fimbriae of *Escherichia coli*. *PLoS One*. 2013; 8(1):e52835. Epub 2013/02/06. <https://doi.org/10.1371/journal.pone.0052835> PMID: 23382825; PubMed Central PMCID: PMC3559732.
10. Wurpel DJ, Totsika M, Allsopp LP, Webb RI, Moriel DG, Schembri MA. Comparative proteomics of uropathogenic *Escherichia coli* during growth in human urine identify UCA-like (UCL) fimbriae as an adherence factor involved in biofilm formation and binding to uroepithelial cells. *Journal of proteomics*. 2016; 131:177–89. <https://doi.org/10.1016/j.jprot.2015.11.001> PMID: 26546558.
11. Wurpel DJ, Moriel DG, Totsika M, Easton DM, Schembri MA. Comparative analysis of the uropathogenic *Escherichia coli* surface proteome by tandem mass-spectrometry of artificially induced outer membrane vesicles. *Journal of proteomics*. 2015; 115:93–106. <https://doi.org/10.1016/j.jprot.2014.12.005> PMID: 25534882.
12. Spaulding CN, Klein RD, Ruer S, Kau AL, Schreiber HL, Cusumano ZT, et al. Selective depletion of uropathogenic *E. coli* from the gut by a FimH antagonist. *Nature*. 2017; 546(7659):528–32. Epub 2017/06/15. <https://doi.org/10.1038/nature22972> PMID: 28614296; PubMed Central PMCID: PMC5654549.
13. Pellegrino R, Scavone P, Umpierrez A, Maskell DJ, Zunino P. *Proteus mirabilis* uroepithelial cell adhesin (UCA) fimbria plays a role in the colonization of the urinary tract. *Pathogens and Disease*. 2013; 67(2):104–7. <https://doi.org/10.1111/2049-632X.12027> PMID: 23620155.
14. Wray SK, Hull SI, Cook RG, Barrish J, Hull RA. Identification and characterization of a uroepithelial cell adhesin from a uropathogenic isolate of *Proteus mirabilis*. *Infection and immunity*. 1986; 54(1):43–9. <https://doi.org/10.1128/iai.54.1.43-49.1986> PMID: 2875952; PubMed Central PMCID: PMC260114.
15. Buts L, Bouckaert J, De Genst E, Loris R, Oscarson S, Lahmann M, et al. The fimbrial adhesin F17-G of enterotoxigenic *Escherichia coli* has an immunoglobulin-like lectin domain that binds N-acetylglucosamine. *Molecular Microbiology*. 2003; 49(3):705–15. <https://doi.org/10.1046/j.1365-2958.2003.03600.x> PMID: 12864853.
16. Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M, Forde BM, et al. Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111(15):5694–9. Epub 2014/04/08. <https://doi.org/10.1073/pnas.1322678111> PMID: 24706808; PubMed Central PMCID: PMC3992628.
17. Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, Khanh Nhu NT, Roberts LW, Stanton-Cook M, et al. Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of *Escherichia coli* ST131. *MBio*. 2016; 7(2):e00347–16. <https://doi.org/10.1128/mBio.00347-16> PMID: 27118589; PubMed Central PMCID: PMC4850260.
18. Nyerges A, Csorgo B, Nagy I, Balint B, Bihari P, Lazar V, et al. A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proceedings of the National Academy of Sciences of the United States of America*. 2016; 113(9):2502–7. <https://doi.org/10.1073/pnas.1520040113> PMID: 26884157; PubMed Central PMCID: PMC4780621.

19. Munch R, Hiller K, Grote A, Scheer M, Klein J, Schobert M, et al. Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. *Bioinformatics*. 2005; 21(22):4187–9. <https://doi.org/10.1093/bioinformatics/bti635> PMID: 16109747.
20. Seo SW, Kim D, Szubin R, Palsson BO. Genome-wide Reconstruction of OxyR and SoxRS Transcriptional Regulatory Networks under Oxidative Stress in *Escherichia coli* K-12 MG1655. *Cell reports*. 2015; 12(8):1289–99. <https://doi.org/10.1016/j.celrep.2015.07.043> PMID: 26279566.
21. Jiang W, Ubhayasekera W, Pearson MM, Knight SD. Structures of two fimbrial adhesins, AtfE and UcaD, from the uropathogen *Proteus mirabilis*. *Acta Crystallographica Section D*. 2018; 74(Pt 11):1053–62. <https://doi.org/10.1107/S2059798318012391> PMID: 30387764.
22. Holm L, Laakso LM. Dali server update. *Nucleic acids research*. 2016; 44(W1):W351–5. Epub 2016/05/01. <https://doi.org/10.1093/nar/gkw357> PMID: 27131377; PubMed Central PMCID: PMC4987910.
23. Spiwok V. CH/pi Interactions in Carbohydrate Recognition. *Molecules*. 2017; 22(7). <https://doi.org/10.3390/molecules22071038> PMID: 28644385; PubMed Central PMCID: PMC6152320.
24. Peek AS, Souza V, Eguiarte LE, Gaut BS. The interaction of protein structure, selection, and recombination on the evolution of the type-1 fimbrial major subunit (fimA) from *Escherichia coli*. *Journal of molecular evolution*. 2001; 52(2):193–204. <https://doi.org/10.1007/s002390010148> PMID: 11244580.
25. Boyd EF, Hartl DL. Diversifying selection governs sequence polymorphism in the major adhesin proteins fimA, papA, and sfaA of *Escherichia coli*. *Journal of molecular evolution*. 1998; 47(3):258–67. <https://doi.org/10.1007/pl00006383> PMID: 9732452
26. Sokurenko EV, Chesnokova V, Dykhuizen DE, Ofek I, Wu XR, Krogfelt KA, et al. Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. *Proceedings of the National Academy of Sciences of the United States of America*. 1998; 95(15):8922–6. <https://doi.org/10.1073/pnas.95.15.8922> PMID: 9671780; PubMed Central PMCID: PMC21178.
27. Schwartz DJ, Kalas V, Pinkner JS, Chen SL, Spaulding CN, Dodson KW, et al. Positively selected FimH residues enhance virulence during urinary tract infection by altering FimH conformation. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110(39):15530–7. <https://doi.org/10.1073/pnas.1315203110> PMID: 24003161; PubMed Central PMCID: PMC3785778.
28. Hammarlof DL, Kroger C, Owen SV, Canals R, Lacharme-Lora L, Wenner N, et al. Role of a single non-coding nucleotide in the evolution of an epidemic African clade of *Salmonella*. *Proceedings of the National Academy of Sciences of the United States of America*. 2018; 115(11):E2614–E23. <https://doi.org/10.1073/pnas.1714718115> PMID: 29487214; PubMed Central PMCID: PMC5856525.
29. Storz G, Imlay JA. Oxidative stress. *Current opinion in microbiology*. 1999; 2(2):188–94. [https://doi.org/10.1016/s1369-5274\(99\)80033-2](https://doi.org/10.1016/s1369-5274(99)80033-2) PMID: 10322176.
30. Shanks RM, Stella NA, Kalivoda EJ, Doe MR, O'Dee DM, Lathrop KL, et al. A *Serratia marcescens* OxyR homolog mediates surface attachment and biofilm formation. *J Bacteriol*. 2007; 189(20):7262–72. Epub 2007/08/07. <https://doi.org/10.1128/JB.00859-07> PMID: 17675374; PubMed Central PMCID: PMC2168423.
31. Wu J, Lin X, Xie H. OxyR is involved in coordinate regulation of expression of fimA and sod genes in *Porphyromonas gingivalis*. *FEMS Microbiol Lett*. 2008; 282(2):188–95. Epub 2008/03/22. <https://doi.org/10.1111/j.1574-6968.2008.01116.x> PMID: 18355277; PubMed Central PMCID: PMC2430334.
32. Rychlik I, Barrow PA. *Salmonella* stress management and its relevance to behaviour during intestinal colonisation and infection. *EMS Microbiology Reviews*. 2005; 29(5):1021–40. <https://doi.org/10.1016/j.femsre.2005.03.005> PMID: 16023758.
33. Sistrunk JR, Nickerson KP, Chanin RB, Rasko DA, Faherty CS. Survival of the Fittest: How Bacterial Pathogens Utilize Bile To Enhance Infection. *Clinical microbiology rev*. 2016; 29(4):819–36. <https://doi.org/10.1128/CMR.00031-16> PMID: 27464994; PubMed Central PMCID: PMC5010752.
34. Prieto AI, Ramos-Morales F, Casadesus J. Repair of DNA damage induced by bile salts in *Salmonella enterica*. *Genetics*. 2006; 174(2):575–84. <https://doi.org/10.1534/genetics.106.060889> PMID: 16888329; PubMed Central PMCID: PMC1602091.
35. Lee KK, Harrison BA, Latta R, Altman E. The binding of *Proteus mirabilis* nonagglutinating fimbriae to ganglio-series asialoglycolipids and lactosyl ceramide. *Canadian journal of microbiology*. 2000; 46(10):961–6. PMID: 11068685.
36. Smith DF, Prieto PA, McCrumb DK, Wang WC. A novel sialylfucopentaose in human milk. Presence of this oligosaccharide is not dependent on expression of the secretor or Lewis fucosyltransferases. *Journal of Biological Chemistry*. 1987; 262(25):12040–7. PMID: 3624247.
37. Forde BM, Roberts LW, Phan MD, Peters KM, Fleming BA, Russell CW, et al. Population dynamics of an *Escherichia coli* ST131 lineage during recurrent urinary tract infection. *Nature communications*. 2019; 10(1):3643. <https://doi.org/10.1038/s41467-019-11571-5> PMID: 31409795; PubMed Central PMCID: PMC6692316.

38. Kalas V, Hibbing ME, Maddirala AR, Chugani R, Pinkner JS, Mydock-McGrane LK, et al. Structure-based discovery of glycomimetic FmlH ligands as inhibitors of bacterial adhesion during urinary tract infection. *Proceedings of the National Academy of Sciences of the United States of America*. 2018; 115(12):E2819–E28. <https://doi.org/10.1073/pnas.1720140115> PMID: 29507247; PubMed Central PMCID: PMC5866590.
39. Phan MD, Peters KM, Sarkar S, Lukowski SW, Allsopp LP, Gomes Moriel D, et al. The serum resistome of a globally disseminated multidrug resistant uropathogenic *Escherichia coli* clone. *PLoS Genet*. 2013; 9(10):e1003834. <https://doi.org/10.1371/journal.pgen.1003834> PMID: 24098145; PubMed Central PMCID: PMC3789825.
40. Stocks CJ, Phan M-D, Achard MES, Nhu NTK, Condon ND, Gawthorne JA, et al. Uropathogenic *Escherichia coli* employs both evasion and resistance to subvert innate immune-mediated zinc toxicity for dissemination. *Proc Natl Acad Sci USA*. 2019; 116(13):6341–50. Epub Epub 2019 Mar 7. <https://doi.org/10.1073/pnas.1820870116> PMID: 30846555
41. Phan MD, Peters KM, Alvarez Fraga L, Wallis SC, Hancock SJ, Nhu NTK, et al. Plasmid-mediated ciprofloxacin resistance imparts a selective advantage on *Escherichia coli* ST131. *Antimicrob Agents Chemother*. 2022; 66(1):e0214621. Epub 2021/11/16. <https://doi.org/10.1128/AAC.02146-21> PMID: 34780264; PubMed Central PMCID: PMC8765324.
42. Nhu NTK, Phan MD, Forde BM, Murthy AMV, Peters KM, Day CJ, et al. Complex multilevel control of hemolysin production by uropathogenic *Escherichia coli*. *mBio*. 2019; 10(5):e02248–19. Epub 2019/10/03. <https://doi.org/10.1128/mBio.02248-19> PMID: 31575773; PubMed Central PMCID: PMC6775461.
43. Nhu NTK, Phan MD, Peters KM, Lo AW, Forde BM, Min Chong T, et al. Discovery of new genes involved in curli production by a uropathogenic *Escherichia coli* strain from the highly virulent O45:K1:H7 lineage. *mBio*. 2018; 9(4):pii: e01462-18. Epub 2018/08/23. <https://doi.org/10.1128/mBio.01462-18> PMID: 30131362; PubMed Central PMCID: PMC6106082.
44. Hancock SJ, Phan MD, Luo Z, Lo AW, Peters KM, Nhu NTK, et al. Comprehensive analysis of IncC plasmid conjugation identifies a crucial role for the transcriptional regulator AcaB. *Nat Microbiol*. 2020; 5(11):1340–8. Epub 2020/08/19. <https://doi.org/10.1038/s41564-020-0775-0> PMID: 32807890.
45. Russell CW, Fleming BA, Jost CA, Tran A, Stenquist AT, Wambaugh MA, et al. Context-Dependent Requirements for FimH and Other Canonical Virulence Factors in Gut Colonization by Extraintestinal Pathogenic *Escherichia coli*. *Infection and immunity*. 2018; 86(3). <https://doi.org/10.1128/IAI.00746-17> PMID: 29311232; PubMed Central PMCID: PMC5820936.
46. Furlong EJ, Lo AW, Kurth F, Premkumar L, Totsika M, Achard MES, et al. A shape-shifting redox foldase contributes to *Proteus mirabilis* copper resistance. *Nature communications*. 2017; 8:16065. <https://doi.org/10.1038/ncomms16065> PMID: 28722010; PubMed Central PMCID: PMC5524982.
47. Shewell LK, Day CJ, Jen FE, Haselhorst T, Atack JM, Reijneveld JF, et al. All major cholesterol-dependent cytolysins use glycans as cellular receptors. *Science advances*. 2020; 6(21):eaaz4926. <https://doi.org/10.1126/sciadv.aaz4926> PMID: 32494740; PubMed Central PMCID: PMC7244308.
48. Waespy M, Gbem TT, Elenschneider L, Jeck AP, Day CJ, Hartley-Tassell L, et al. Carbohydrate Recognition Specificity of Trans-sialidase Lectin Domain from *Trypanosoma congolense*. *PLoS neglected tropical diseases*. 2015; 9(10):e0004120. <https://doi.org/10.1371/journal.pntd.0004120> PMID: 26474304; PubMed Central PMCID: PMC4608562.