

ARTICLE

<https://doi.org/10.1038/s41467-019-09814-6>

OPEN

# Unique structural features of a bacterial autotransporter adhesin suggest mechanisms for interaction with host macromolecules

Jason J. Paxman <sup>1,9</sup>, Alvin W. Lo<sup>2,9</sup>, Matthew J. Sullivan<sup>3</sup>, Santosh Panjikar <sup>4,5</sup>, Michael Kuiper<sup>6</sup>, Andrew E. Whitten <sup>7</sup>, Geqing Wang<sup>1</sup>, Chi-Hao Luan<sup>8</sup>, Danilo G. Moriel<sup>2</sup>, Lendl Tan <sup>2</sup>, Kate M. Peters<sup>2</sup>, Minh-Duy Phan <sup>2</sup>, Christine L. Gee <sup>4</sup>, Glen C. Ulett <sup>3</sup>, Mark A. Schembri <sup>2,10</sup> & Begoña Heras <sup>1,10</sup>

Autotransporters are the largest family of outer membrane and secreted proteins in Gram-negative bacteria. Most autotransporters are localised to the bacterial surface where they promote colonisation of host epithelial surfaces. Here we present the crystal structure of UpaB, an autotransporter that is known to contribute to uropathogenic *E. coli* (UPEC) colonisation of the urinary tract. We provide evidence that UpaB can interact with glycosaminoglycans and host fibronectin. Unique modifications to its core  $\beta$ -helical structure create a groove on one side of the protein for interaction with glycosaminoglycans, while the opposite face can bind fibronectin. Our findings reveal far greater diversity in the autotransporter  $\beta$ -helix than previously thought, and suggest that this domain can interact with host macromolecules. The relevance of these interactions during infection remains unclear.

<sup>1</sup>Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne 3086 VIC, Australia. <sup>2</sup>School of Chemistry and Molecular Biosciences, and Australian Infectious Diseases Research Centre, The University of Queensland, Brisbane 4072 QLD, Australia. <sup>3</sup>School of Medical Science, and Menzies Health Institute Queensland, Griffith University, Gold Coast 4222 QLD, Australia. <sup>4</sup>Macromolecular Crystallography, Australian Synchrotron, Clayton 3168 VIC, Australia. <sup>5</sup>Department of Molecular Biology and Biochemistry, Monash University, Melbourne 3800 VIC, Australia. <sup>6</sup>Molecular & Materials Modelling group Data61, CSIRO, Docklands, Melbourne 8012 VIC, Australia. <sup>7</sup>Australian Centre for Neutron Scattering, Australian Nuclear Science and Technology Organisation, Lucas Heights 2234 NSW, Australia. <sup>8</sup>High Throughput Analysis Laboratory and Department of Molecular Biosciences, Northwestern University, Chicago 60208 IL, USA. <sup>9</sup>These authors contributed equally: Jason J. Paxman, Alvin W. Lo. <sup>10</sup>These authors jointly supervised this work: Mark A. Schembri, Begoña Heras. Correspondence and requests for materials should be addressed to M.A.S. (email: [m.schembri@uq.edu.au](mailto:m.schembri@uq.edu.au)) or to B.H. (email: [b.heras@latrobe.edu.au](mailto:b.heras@latrobe.edu.au))

**A** key factor in the establishment of infections by most bacterial pathogens is their adherence to host epithelial cells<sup>1,2</sup>. Autotransporters (ATs) are the largest group of outer membrane and secreted proteins in bacteria and play important roles in virulence, including promoting adhesion<sup>3</sup>. ATs share a common domain organisation consisting of a Sec-dependent signal sequence, a passenger or  $\alpha$ -domain and a C-terminal translocator  $\beta$ -domain<sup>4</sup>. The signal sequence and  $\beta$ -domain are required for transport of the  $\alpha$ -domain through the inner and outer membranes, respectively. The  $\alpha$ -domain is the functional portion of the protein and can drive phenotypes including cytotoxicity, aggregation, adhesion and/or invasion, features that enhance bacterial virulence, colonisation, biofilm formation, persistence and resistance to host innate defence mechanisms<sup>3–5</sup>. In an era of increasing antimicrobial resistance, a greater understanding of the mechanisms by which ATs augment bacterial pathogenesis is required if we are to develop new strategies to combat infections caused by multidrug-resistant pathogens.

Despite the abundance of AT genes in the GenBank sequence database and the importance of these proteins in bacterial pathogenesis, structural information about AT  $\alpha$ -domains remains limited. To date, only 11  $\alpha$ -domains and some small fragments of trimeric ATs (a separate subfamily of ATs that are obligate trimers<sup>6–8</sup>) have been structurally characterised, and consequently, AT molecular mechanisms of action are largely unknown.

Based on limited structural information, AT  $\alpha$ -domains adopt a general architecture comprising a long narrow right-handed  $\beta$ -helix that is embellished with loops and/or other small domains (e.g. trypsin-like domains) that confer different functional properties<sup>9–12</sup>. One of the best characterised ATs is the antigen 43 (Ag43) protein, which belongs to the largest and most diverse AT subfamily, the AIDA-I type class<sup>13</sup>. Ag43 is the only member with a known mechanism of action, whereby a head to tail dimerisation of Ag43 monomers on the surface of adjacent bacterial cells promotes aggregation and biofilm formation via a molecular ‘velcro-like’ mechanism<sup>14–17</sup>.

The surface-exposed  $\alpha$ -domains of different AIDA-I type ATs exhibit extensive sequence variation<sup>13</sup>. An example of this is represented by UpaB from uropathogenic *Escherichia coli* (UPEC), the major aetiologic agent of urinary tract infection and a primary cause of sepsis<sup>18</sup>. UpaB is an AT that mediates UPEC adherence to extracellular matrix (ECM) proteins and enhances UPEC colonisation of the urinary tract<sup>19</sup>. Here we determine the structure of UpaB at high resolution and reveal that it adopts a unique architecture potentially comprising two distinct binding sites. One binding site is formed by significant extensions to its  $\beta$ -helix that form a groove that can interact with glycosaminoglycans. On its opposite face, a second binding region can interact with human fibronectin (FN) type III (FnIII). Our results suggest that the AT  $\beta$ -helix may have diverse roles in addition to acting as a structural scaffold<sup>5</sup>.

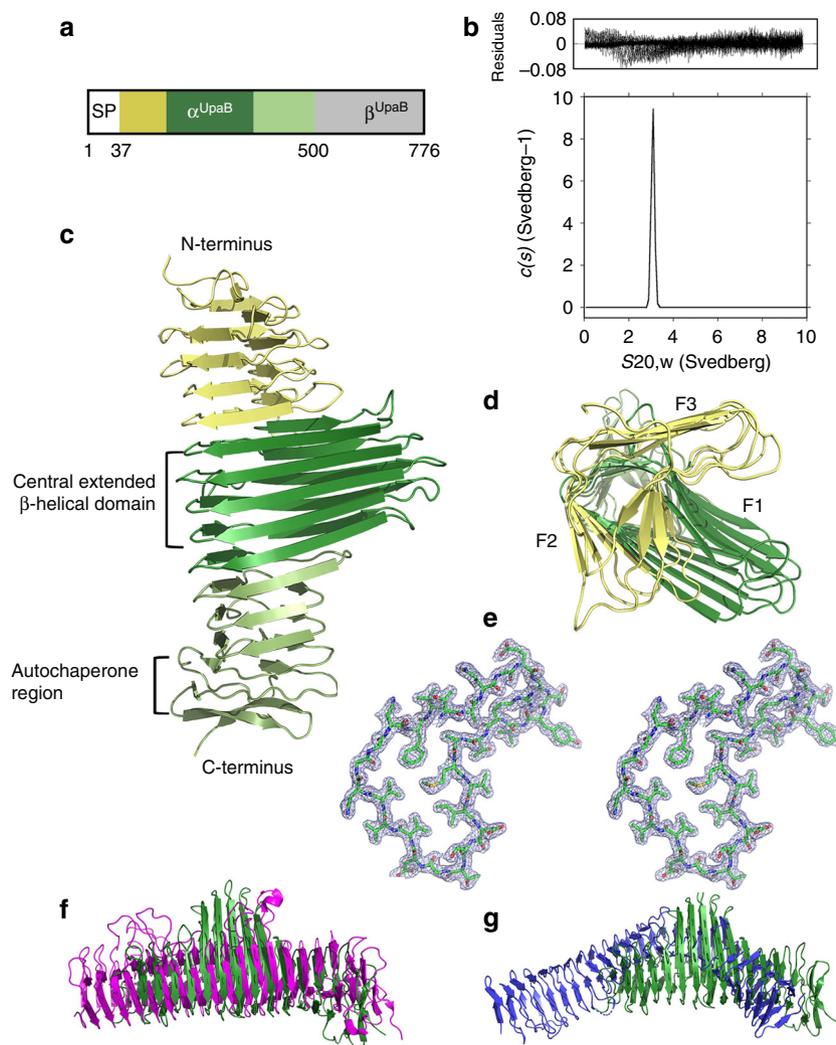
## Results

**Characterisation of UpaB reveals that it does not exhibit self-association properties.** AIDA-I-type ATs including Ag43, TibA and AIDA-I belong to a group of self-associating ATs that promote bacterial aggregation and biofilm formation<sup>20</sup>. In the case of Ag43, aggregation is mediated via a series of hydrogen bonds, hydrophobic interactions and Van der Waals forces that drive a head-to-tail dimerisation of Ag43 monomers on the surface of adjacent cells<sup>17</sup>. We previously showed that UpaB from the UPEC reference strain CFT073 does not mediate cell aggregation when overexpressed in *E. coli* laboratory strains<sup>19</sup>, but under certain

conditions, it may have a slight indirect effect on these phenotypes<sup>21</sup>. UpaB is composed of an N-terminal signal sequence (residues 1–37), an  $\alpha$ -domain (residues 38–500) and a  $\beta$ -domain (residues 501–776) (Fig. 1a). In order to understand the functional properties of UpaB, we cloned, expressed and purified the region encoding the UpaB  $\alpha$ -domain ( $\alpha^{\text{UpaB}}$ ) from UPEC CFT073 and used analytical ultracentrifugation sedimentation velocity experiments to assess its propensity to self-associate in solution. At 0.5, 1 and 2.2 mg ml<sup>-1</sup>,  $\alpha^{\text{UpaB}}$  produced a single sedimentation boundary and a continuous sedimentation-coefficient distribution ( $c(s)$ ) (Fig. 1b) to give a single species with a standardised sedimentation coefficient of 3.1 s. Analysis by continuous mass distribution ( $c(M)$ ) gave a molecular weight of approximately 52.7 kDa, consistent with a monomeric species. Likewise, small-angle X-ray scattering (SAXS) of  $\alpha^{\text{UpaB}}$  at concentrations <2.7 mg ml<sup>-1</sup> was consistent with a monodisperse protein population (Supplementary Fig. 1), with the estimated mass, radius of gyration ( $R_g$ ) and the maximum linear dimension ( $D_{\text{max}}$ ) from the experimental pair-distance distribution profile ( $p(r)$ ) yielding values close to what would be expected for a solution of monomeric  $\alpha^{\text{UpaB}}$  (Supplementary Table 1). Thus, unlike the  $\alpha$ -domain of Ag43 ( $\alpha^{\text{Ag43}}$ ), recombinant  $\alpha^{\text{UpaB}}$  does not self-associate. This fundamental difference in the functional properties of both proteins led us to determine the crystal structure of  $\alpha^{\text{UpaB}}$ .

**Overall UpaB structure.** The  $\alpha^{\text{UpaB}}$  crystal structure was solved to 1.9 Å resolution (crystallographic  $R_{\text{factor}}$  of 17.5%;  $R_{\text{free}}$  21.8%). The crystallographic refinement statistics are reported in Table 1. The structure was solved from a xenon derivative by single isomorphous replacement with anomalous scattering. One molecule of  $\alpha^{\text{UpaB}}$  was found in the asymmetric unit. The crystal structure of  $\alpha^{\text{UpaB}}$  exhibited a right-handed three-stranded  $\beta$ -helix consisting of 13 turns (Fig. 1c), with each triangular turn containing three faces, F1, F2 and F3 faces (Fig. 1d). The  $\beta$ -helix is predominantly stabilised by an inter-strand network of hydrogen bonds. The interior of the  $\beta$ -helix is packed mostly by long stacks of aliphatic residues (Fig. 1e), whereas the exterior is largely acidic in nature. At the C-terminus of the  $\beta$ -helix,  $\alpha^{\text{UpaB}}$  forms a two-stranded  $\beta$ -sandwich that is capped by a three-stranded  $\beta$ -meander motif. This latter region resembles the autochaperone region that is required for the folding of AT  $\alpha$ -domains on the cell surface<sup>22</sup>. A DALI structural alignment revealed that  $\alpha^{\text{UpaB}}$  shared only a low resemblance to other AT structures in the PDB, with highest similarity to the *Bordetella pertussis* pertactin AT (PDB 1DAB); (18% sequence identity, Z-score 24.5 and r.m.s.d of 2.8 Å between 358 equivalent Ca atoms) (Fig. 1f). Compared to  $\alpha^{\text{Ag43}}$ ,  $\alpha^{\text{UpaB}}$  shares only 12% sequence identity, Z-score 14.6 and r.m.s.d of 3.8 Å between 242 equivalent Ca atoms (Fig. 1g). Further comparison with other ATs<sup>9–12,17</sup> revealed the  $\alpha^{\text{UpaB}}$   $\beta$ -helix is wider and shorter with a total length of 75 Å. However, the most striking feature of the  $\alpha^{\text{UpaB}}$   $\beta$ -helix is the extended  $\beta$ -strands within turns 6–10, which reach up to 32 Å in length. In addition, the strands linking turns 2–3, 3–4, 4–5 and 5–6 are lengthened into a consecutive series of large loops. To date, these structural features have not been observed in the  $\alpha$ -domain of any other AT protein<sup>23</sup>.

**UpaB can bind glycosaminoglycans.** The  $\beta$ -strand extensions contributed by turns 6–10 and the long loops protruding between turns 2–6 form a long hydrophilic groove 11 Å wide and 12.5 Å deep on the F1 face of  $\alpha^{\text{UpaB}}$  (Fig. 2a, b). Sidechains from E165, S188, N189, Q197, T230 and E293 protrude into the groove and largely determine its slightly acidic nature. The results of the DALI search using  $\alpha^{\text{UpaB}}$  were further analysed to define a role



**Fig. 1** The structure of the UpaB functional  $\alpha$ -domain ( $\alpha^{\text{UpaB}}$ ). **a** Domain organisation of UpaB comprising an N-terminal signal sequence (SP; residues 1–37), an  $\alpha$ -domain ( $\alpha^{\text{UpaB}}$ ; residues 38–500) and a  $\beta$ -domain ( $\beta^{\text{UpaB}}$ ; residues 501–776). **b** Analytical ultracentrifugation sedimentation velocity analysis of  $\alpha^{\text{UpaB}}$ . The continuous standardised sedimentation distribution [ $c(s)$ ] shows that UpaB at  $2.2 \text{ mg ml}^{-1}$  exists as a  $3.1 s_{20,w}$  monomer. **c** Cartoon representation of the  $\alpha^{\text{UpaB}}$  structure, including **d** top view. The central domain consisting of extended  $\beta$ -strands is shown in dark green. The N-terminal and C-terminal  $\beta$ -helical domains are shown in yellow and light green, respectively. The top view has F1, F2 and F3 faces shown. **e** Stereo view of the  $2F_o - F_c$  electron density map contoured at  $1\sigma$  of the cross-section of the  $\alpha^{\text{UpaB}}$   $\beta$ -helix. Structural comparison of UpaB (green) with the  $\alpha$ -domain of **f** pertactin (from *B. pertussis*; magenta; PDB 1DAB) and **g** Ag43a (from UPEC; blue; PDB 4KH3)

for its groove and revealed that UpaB shared low structural similarity to polysaccharide degrading enzymes (1BHE, 5GKD, 4C2L), which are also composed of a  $\beta$ -helix with a prominent groove for binding polysaccharides<sup>24–26</sup>. The  $\alpha^{\text{UpaB}}$  groove most closely resembled the glycosaminoglycan (GAG) lyase chondroitinase B (PDB 1OFL) from *Pedobacter heparinus*<sup>27</sup> (8% sequence identity, Z-score 16.5 and r.m.s.d of  $3.3 \text{ \AA}$ ) (Fig. 2c). Chondroitinase B is the closest homolog known to interact with human polysaccharides. Importantly,  $\alpha^{\text{UpaB}}$  shares a putative active site with chondroitinase B and other GAG lyases, located just outside the groove (Fig. 2c). This site comprises UpaB Lys 256 and Lys 343 situated in similar positions to chondroitinase B Lys250/Arg271 Brønsted base/acid pair required to break the glycosidic bonds of GAGs<sup>28</sup>. In chondroitinase B and other GAG lyases, the Lys250/Arg271 would be situated proximal to a bound calcium ion required for neutralisation of the GAG carboxylic group during bond cleavage. Indeed, we identified electron density associated with the  $\alpha^{\text{UpaB}}$  lysine pair likely to be a bound calcium (Supplementary Fig. 2). Similar to other lyases,

this calcium ion would be held in place by the neighbouring  $\alpha^{\text{UpaB}}$  Glu 314 and Asn 316 residues. The likelihood of a GAG binding within the UpaB groove was tested using docking simulations (Fig. 2a). A model of a GAG was constructed and docked into the  $\alpha^{\text{UpaB}}$  groove using Autodock Vina. All of the docking conformations showed an interaction with the  $\alpha^{\text{UpaB}}$  groove, with one of the top conformations displaying an interaction with the putative lyase active site resembling a pre-cleavage state (Supplementary Fig. 2b). This binding conformation exhibited a significant predicted binding affinity of  $-9.4 \text{ kcal mol}^{-1}$  (free energy of binding), based on an extensive hydrogen bonding network between the GAG hydroxyl groups and a number of polar residues within and around the  $\alpha^{\text{UpaB}}$  groove.

This investigation was followed by the screening of  $\alpha^{\text{UpaB}}$  against 2788 compounds (including 88 carbohydrate molecules) in a fluorescence thermal shift-based assay (Fig. 2d). Significant binding was shown to two ‘GAG-like’ molecules, Tn Antigen GalN- $\alpha$ 1-O-Ser and lacto-N-neohexaose (Fig. 2d). GalN- $\alpha$ 1-O-Ser closely resembles the O-glycosidic-linked saccharide to serine

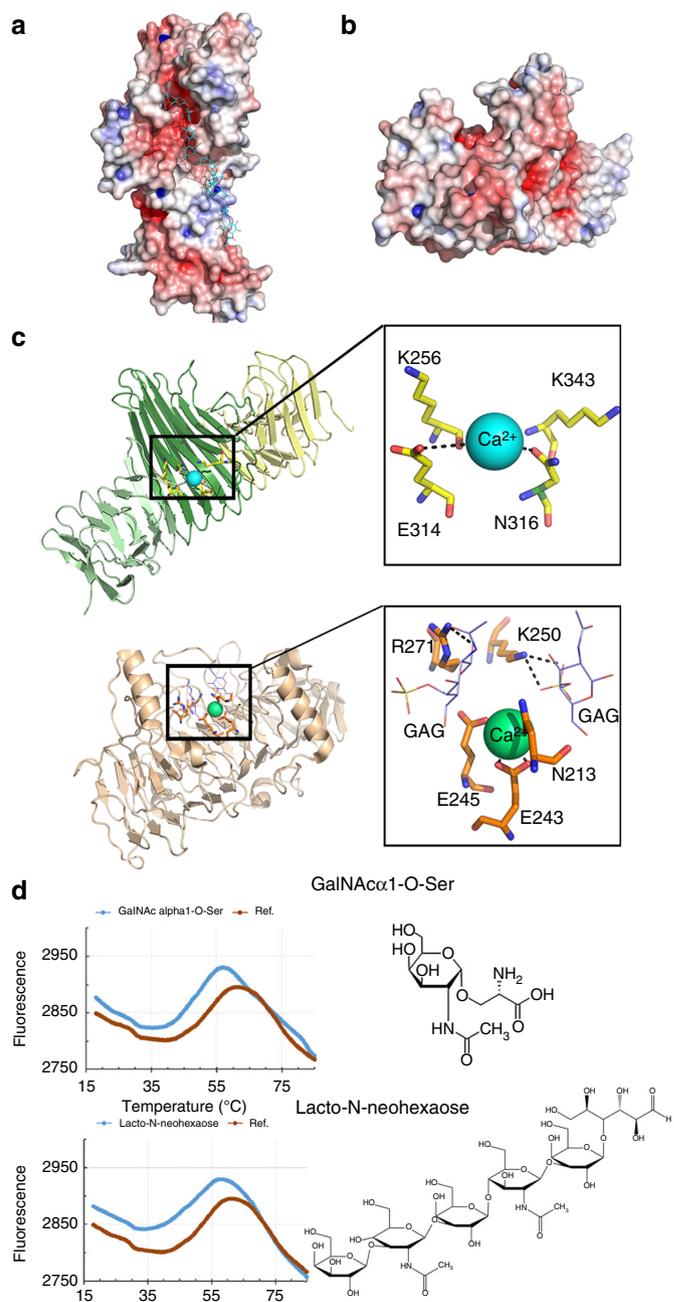
**Table 1** UpaB data collection and statistics

	Native	Xenon
Data collection		
Resolution (Å)	50.0-1.97 (2.04-1.97)	48.57-2.50 (2.64-2.50)
Wavelength (Å)	0.9537	1.3776
Space group	P3 <sub>1</sub> 21	P3 <sub>1</sub> 21
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	68.6, 68.6, 165.6	69.2, 69.2, 166.0
$\alpha$ , $\beta$ , $\gamma$ (degrees)	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Molecules per ASU	1	1
Total no. of reflections	243,523	695,483
No. of unique reflections	32,914 (3219)	16,619 (2351)
Completeness (%)	99.6 (99.6)	100.0 (100.0)
Redundancy	7.4 (6.8)	41.8 (42.9)
<i>I</i> / $\sigma$ ( <i>I</i> )	13.5 (2.7)	30.3 (9.2)
<i>R</i> <sub>merge</sub>	7.8 (71.0)	12.3 (56.0)
Phasing		
Resolution		2.5
No. of sites		8
Figure of merit		0.52
Refinement		
Resolution (Å)	33.96-1.97 (2.04-1.97)	
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	17.5/21.8 (0.217/0.245)	
No. of reflections	32,800 (3208)	
No. of atoms		
Protein	3158	
Water	240	
<i>B</i> -factors (Å <sup>2</sup> )		
Protein	29.92	
Water	40.13	
RMS deviations		
Bond lengths (Å)	0.007	
Bond angles (degrees)	0.92	

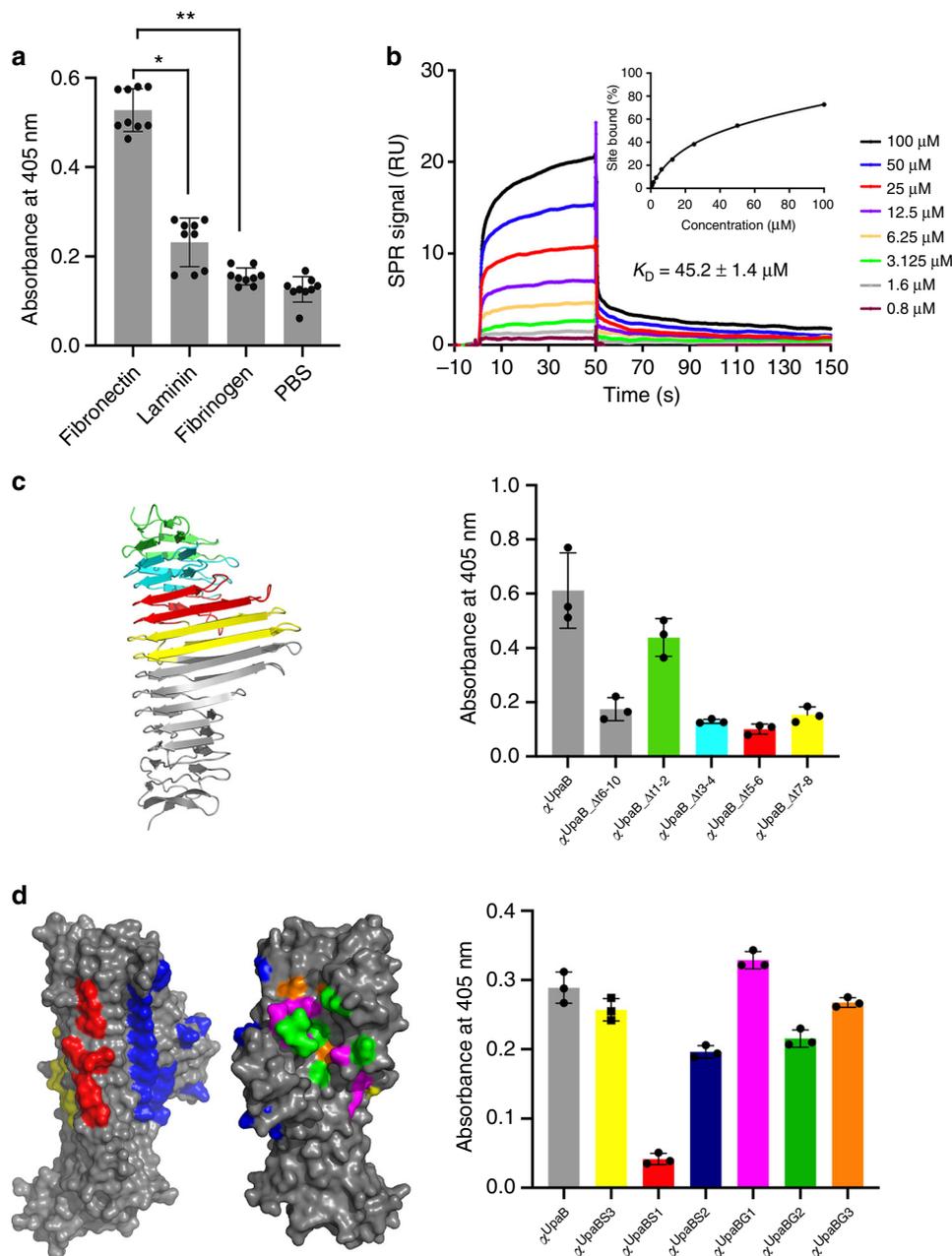
Values in parentheses refer to the highest resolution shell

complex that anchors most GAGs to their core proteins, and the lacto-*N*-neohexaose is representative of a main chain GAG<sup>29</sup>. The role of the UpaB groove in this binding was shown by repeating this assay with a UpaB mutant ( $\alpha$ UpaB<sub>G1</sub>), designed by alanine substitutions to the prominent residues that stabilise the GAG interaction identified in our molecular docking studies (E165A, N189A, Q197A, N200A, Q203A, K256A and N316A). Although we confirmed that these alterations did not affect the secondary structure of  $\alpha$ UpaB<sub>G1</sub> (Supplementary Fig. 3a) and  $\alpha$ UpaB<sub>G1</sub> behaved in solution similar to the native protein (Supplementary Fig. 3b), this mutant was unable to bind the ‘GAG-like’ molecules as shown by the overlapping melting curve plots of  $\alpha$ UpaB<sub>G1</sub> in the presence and absence of the GAGs (Supplementary Fig. 3c). Further analyses revealed that  $\alpha$ UpaB did not display a broad affinity for some common GAGs found in the urinary tract including chondroitin sulfate A, B, C and heparin sulfate<sup>29,30</sup> (Supplementary Figs. 4 and 5).

Overall, these results show that  $\alpha$ UpaB can bind GAG-like molecules and that this binding is lost when we mutate the GAG-binding site. Our data also indicate that  $\alpha$ UpaB may display a considerable substrate specificity (chemical library screening identified only two ‘GAG-like’ molecules out of 2788 compounds and we did not observe binding to four polysaccharides found in the urinary tract) and therefore may interact with a limited range of GAGs that we are yet to identify.



**Fig. 2** UpaB can bind glycosaminoglycans. Surface representation of **a**  $\alpha$ UpaB and **b** top view of  $\alpha$ UpaB, with electrostatic potential coloured from the most negative (red) to positive (blue), with a range of  $\pm 10$   $kT e^{-1}$ . The  $\beta$ -strand extensions contributed by turns 6-10 and long loops protruding from between turns 2-6 form an acidic groove. A GAG was modelled into the  $\alpha$ UpaB groove showing that this feature can both accommodate a GAG molecule and place it in proximity to the putative lyase active site. **c** Structural comparison of  $\alpha$ UpaB (green) to *P. heparinus* chondroitinase B (wheat; PDB 1DBG). UpaB shares a  $\beta$ -helix structure, groove, bound calcium (cyan and green) and location of a putative lyase active site with chondroitinase B. UpaB has a putative GAG lyase active site (top right panel) consisting of 256 K and 343 K in proximity to a bound calcium (cyan and green) similar to that of chondroitinase B (lower right panel). **d** Melting curve plots showing the fluorescence intensity (relative fluorescence units (RFU)) of Sypro orange as a function of temperature for purified  $\alpha$ UpaB in the presence of GalN- $\alpha$ 1-O-Ser and Lacto-*N*-neohexaose. The addition of these compounds resulted in a  $T_m$  shift of  $-3.23$  and  $-3.67$  °C, respectively (mean  $T_m$  shift of the 88 carbohydrates screened was  $<0.7$  °C)



**Fig. 3** Functional analysis of the UpaB fibronectin-binding site. **a** Assessment of UpaB binding to human fibronectin, laminin and fibrinogen by enzyme-linked immunosorbent assay (ELISA) using a UpaB-specific polyclonal antibody. UpaB showed highest affinity towards fibronectin. Statistical significance was determined by unpaired two-sample *t* test,  $*P < 0.001$ ,  $n = 9$ ;  $**P < 0.001$ ,  $n = 9$ . **b** Surface plasmon resonance analysis of  $\alpha\text{UpaB}$  binding to immobilised fibronectin. A series of concentrations (0.8–100  $\mu\text{M}$ ) of  $\alpha\text{UpaB}$ , as indicated in the sensogram, were injected over fibronectin. The apparent equilibrium dissociation constant  $K_D$  was determined using a steady-state affinity model. The data are expressed as mean  $\pm$  standard error of the mean (SEM) of three replicates. **c** Assessment of binding to fibronectin by UpaB deletion mutants;  $\alpha\text{UpaB}_{\Delta 6-10}$  (grey);  $\alpha\text{UpaB}_{\Delta 11-2}$  (green),  $\alpha\text{UpaB}_{\Delta 13-4}$  (cyan),  $\alpha\text{UpaB}_{\Delta 15-6}$  (red) and  $\alpha\text{UpaB}_{\Delta 17-8}$  (yellow) using ELISA and a fibronectin-specific polyclonal antibody.  $\alpha\text{UpaB}$  (native) was included as control. Data are shown as the means  $\pm$  standard deviation of three replicates. **d** Assessment of binding to fibronectin by UpaB mutants containing targeted amino acid substitutions using ELISA and a fibronectin-specific polyclonal antibody. Targeted changes were made to various surface features of UpaB including an acidic patch  $\alpha\text{UpaB}_{S1}$  (red; N116A, D119A, N146A, N175A, D217A, K245A, D246A, D281A, R310A and D336A) and polar patch  $\alpha\text{UpaB}_{S2}$  (blue; N110A, K111A, N112A, D142A, N171A, D206A, D208A, N212A, N241A, N274A, N276A, N303A, N305A, K325A, D329A, D331A and D349A) on the F2 face, a hydrophobic patch  $\alpha\text{UpaB}_{S3}$  (yellow; V151A, I221A, V249A, A252G, A253G, Y285A, Y312A and V339A) between the F2 and F3 faces, along with a hydrophobic  $\alpha\text{UpaB}_{G2}$  (green, F101A, Y130A, Y187A, F195A, L201G, L202G, Y260A) and acidic patch  $\alpha\text{UpaB}_{G3}$  (orange, E103A, D138A, E165A, E226A) within the GAG binding groove. Binding to fibronectin by  $\alpha\text{UpaB}_{G1}$  (E165A, N189A, Q197A, N200A, Q203A, K256A and N316A) was also tested. Alteration of the surface acidic patch S1 abolished the ability of UpaB to bind fibronectin. Data are shown as the mean  $\pm$  standard deviation of three replicates

**UpaB contains a FN-binding site.** In addition to GAG containing proteoglycans, the epithelium is comprised of many other glycoproteins including ECM components<sup>31</sup>. *E. coli* expressing UpaB was previously found to bind these ECM proteins<sup>9</sup>. We examined this further by testing the ability of purified  $\alpha^{\text{UpaB}}$  to bind human FN, laminin and fibrinogen (Fig. 3a).  $\alpha^{\text{UpaB}}$  bound strongest to FN, and thus we focussed on understanding this interaction at the molecular level. Using surface plasmon resonance (SPR) we determined a dissociation constant ( $K_D$ ) of  $45.2 \pm 1.4 \mu\text{M}$  between UpaB and FN (Fig. 3b), with the latter immobilised to a CM5 sensor chip using the standard coupling procedure<sup>32</sup>. Although the determined  $K_D$  may be somewhat underestimated owing to the restricted conformational flexibility of the immobilised FN, this binding affinity is consistent with other bacterial FN-binding proteins (including a Fn type III-binding protein)<sup>33–35</sup>. This affinity is also comparable with other physiologically important protein–protein interactions that mediate cell–cell contacts (i.e., T cell receptor–major histocompatibility complex (MHC) complexes range from 2 to  $112 \mu\text{M}$   $K_D$ <sup>36</sup>). Importantly, this binding would not be in the context of a single protein–protein interaction but rather the expression of multiple copies of UpaB on the bacterial cell surface would further enhance the bacterial binding efficiency to FN.

Unlike many other bacterial FN-binding proteins (FnBPs), UpaB does not possess a characteristic GGXXXXV(E/D)(F/I)XX(D/E)T(Xx15)EDT FN-binding repeat (FnBR) sequence<sup>37</sup>. Therefore, to determine the region of  $\alpha^{\text{UpaB}}$  that binds FN, we generated a series of  $\alpha^{\text{UpaB}}$  mutants with specific deletions in  $\beta$ -strand turns, followed by overexpression and purification of the corresponding proteins (Supplementary Fig. 6a). Testing of the mutant proteins for their capacity to bind human FN revealed that deletion of the region encompassing the extended  $\beta$ -strands in turns 6–10 ( $\alpha^{\text{UpaB-}\Delta\text{t6-10}}$ ) resulted in a significant reduction in binding to FN (Fig. 3c). Further analysis involving the progressive deletion of pairs of  $\beta$ -strand turns from the  $\alpha^{\text{UpaB}}$  N-terminus through the extended  $\beta$ -strand region, generating  $\alpha^{\text{UpaB-}\Delta\text{t1-2}}$ ,  $\alpha^{\text{UpaB-}\Delta\text{t3-4}}$ ,  $\alpha^{\text{UpaB-}\Delta\text{t5-6}}$  and  $\alpha^{\text{UpaB-}\Delta\text{t7-8}}$ , demonstrated that the highest loss in FN binding was caused by deletion of turns 3–8. As such, most of the region encompassing the  $\beta$ -strand extensions comprises the primary site for binding FN. The secondary structure of the deletion mutant that was least affected ( $\alpha^{\text{UpaB-}\Delta\text{t1-2}}$ ) and most affected ( $\alpha^{\text{UpaB-}\Delta\text{t5-6}}$ ) to binding FN was confirmed by circular dichroism spectroscopy (Supplementary Fig. 3a). Finally, the same mutations were also generated in the full-length UpaB protein, thereby enabling us to assess its capacity to mediate binding to FN upon translocation to the *E. coli* cell surface. These UpaB deletion mutants, when expressed with their  $\beta$ -domain transporter, were all translocated to the cell surface, as confirmed by whole-cell enzyme-linked immunosorbent assay (ELISA) using polyclonal UpaB antibody (Supplementary Fig. 7a). Subsequent whole-cell ELISA experiments revealed that *E. coli* expressing these UpaB mutant proteins bound to FN in a manner consistent with the results obtained using purified recombinant proteins (Supplementary Fig. 7b).

To determine the specific site within turns 3–8 of  $\alpha^{\text{UpaB}}$  that interacted with FN, we initially investigated the GAG-binding groove on the F1 face, as it was the most prominent structural feature within this region. Utilising our  $\alpha^{\text{UpaB-G1}}$  GAG-binding mutant, along with other mutants containing amino acid substitutions of hydrophobic ( $\alpha^{\text{UpaB-G2}}$ ) and acidic ( $\alpha^{\text{UpaB-G3}}$ ) residues within the groove, we found that these regions had no effect on FN binding as determined by ELISA (Fig. 3d). We next examined the other  $\alpha^{\text{UpaB}}$  faces for possible sites that could bind FN. We made amino acid substitutions to a predominantly acidic

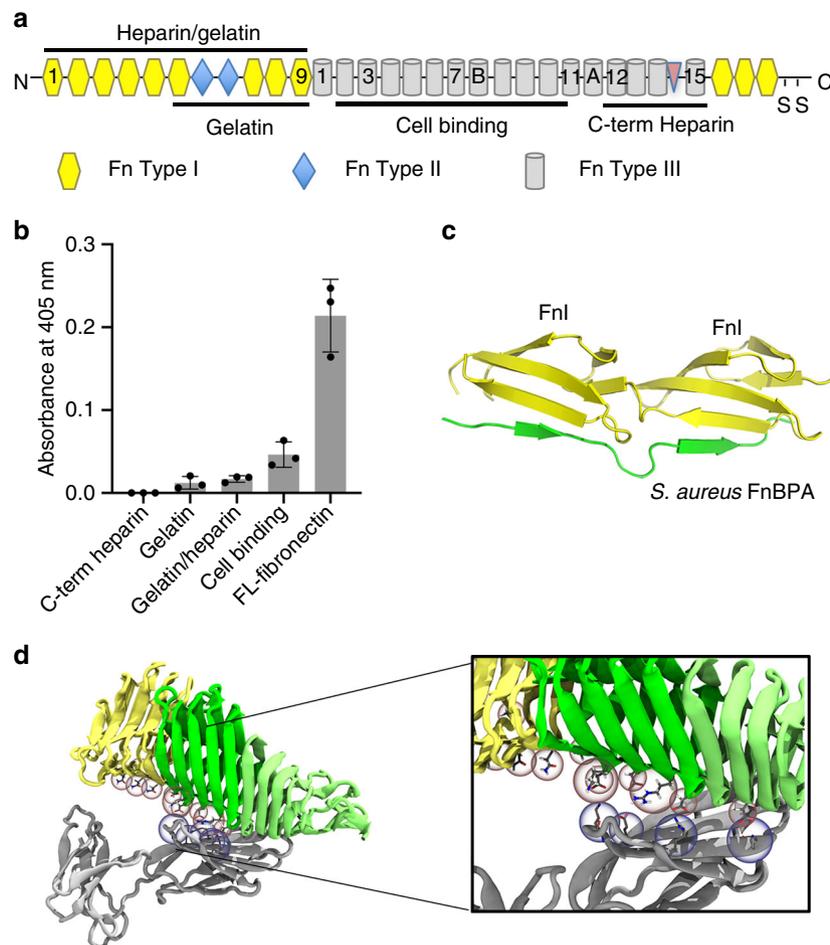
patch ( $\alpha^{\text{UpaB-S1}}$ ) and polar region ( $\alpha^{\text{UpaB-S2}}$ ) on the F2 face and a hydrophobic patch ( $\alpha^{\text{UpaB-S3}}$ ) between the F2 and F3 faces (Fig. 3d).

Substitution of residues D116, D119, N146, N175, D217, K245, D246, D281, R310 and D336 on the F2 face to alanine ( $\alpha^{\text{UpaB-S1}}$ ) caused almost complete loss of FN binding as determined by ELISA, while maintaining the correct secondary structure of  $\alpha^{\text{UpaB-S1}}$  based on circular dichroism spectroscopic analysis (Supplementary Fig. 3a) and displaying a behaviour in solution similar that of the native protein (Supplementary Fig. 3b). This result mapped the FN-binding site to a ladder of charged/polar residues that are contributed from  $\beta$ -strands or loops in consecutive rungs of the  $\alpha^{\text{UpaB}}$   $\beta$ -helix. The only established mode of interaction between FnBPs and FN involves the donation of a series of structurally disordered FnBPs to FN, which upon binding each form an additional  $\beta$ -strand within type I FN modules<sup>37,38</sup>. Thus the  $\alpha^{\text{UpaB}}$ –FN interaction is unique, with interacting residues contributed from already formed  $\beta$ -strands held tightly within the  $\alpha^{\text{UpaB}}$   $\beta$ -helix by a hydrogen-bonding network.

**UpaB can bind FN type III.** In order to further investigate the atypical mode of interaction between UpaB and FN, we determined the region of FN bound by  $\alpha^{\text{UpaB}}$ . FN is composed of 12 type I modules (FnI), 2 type II modules (FnII) and 15–17 type III modules (FnIII)<sup>38</sup>. We obtained commercially available fragments of human FN which included a 45 kDa gelatin-binding fragment (FnI<sub>6–9</sub>, FnII<sub>1–2</sub>), a 70 kDa heparin/gelatin-binding fragment (FnI<sub>1–9</sub>, FnII<sub>1–2</sub>), a 120 kDa cell-binding fragment (FnIII<sub>2–11</sub>) and a 40 kDa C-terminal heparin-binding fragment (FnIII<sub>12–15</sub>)<sup>32</sup> (Fig. 4a, Supplementary Fig. 6b). The binding of  $\alpha^{\text{UpaB}}$  to these FN fragments determined by ELISA revealed that it displayed strongest interaction with the cell binding fragment (FnIII<sub>2–11</sub>) and weak binding to the gelatin (FnI<sub>6–9</sub>, FnII<sub>1–2</sub>) and heparin/gelatin (FnI<sub>1–9</sub>, FnII<sub>1–2</sub>) fragments (Fig. 4b). Given the size of UpaB, this maps its binding site on FN to the first FnIII units in the cell-binding fragment, possibly also including some interaction with the neighbouring FnI units in the gelatin-binding fragment (note that the gelatin [FnI<sub>6–9</sub>, FnII<sub>1–2</sub>] and heparin/gelatin [FnI<sub>1–9</sub>, FnII<sub>1–2</sub>] fragments overlap in this region). The difference observed between UpaB binding to full-length FN compared to its fragments could be attributed to the lack of the FnIII<sub>1</sub> unit within any of the commercially available fragments, which given its location would be an important contribution to the UpaB–Fn interaction. FnIII<sub>1–2</sub> are valid targets for bacterial pathogens such as UPEC, as these 2 units are known to be involved in FN matrix assembly, whereby their disruption could facilitate bacterial spread<sup>39</sup>.

Bacterial interactions with FnIII are uncommon; most of the 100 known bacterial FnBPs bind to the FnI heparin- and gelatin-binding domains via a  $\beta$ -zipper interaction<sup>38</sup> (Fig. 4c). However, there are an increasing number of bacterial proteins including *Haemophilus influenzae* Hap<sup>40</sup>, *Staphylococcus epidermidis* Embp<sup>41</sup>, *Salmonella enterica* serovar Typhimurium ShdA<sup>42</sup> and *Pasteurella multocida* PM1665<sup>32</sup> that bind to FnIII. The AT Hap (13% sequence identity) also primarily binds to FnIII<sub>1–2</sub> via four small binding motifs. These motifs are absent in UpaB and both the UpaB and Hap FN-binding sites do not share the same location.

Overall, the structural basis for how bacterial proteins interact with the FnIII fragment of FN has not been determined, and thus we investigated this interaction using molecular dynamics simulations with our  $\alpha^{\text{UpaB}}$  crystal structure and the structure of human FnIII<sub>1–2</sub> (2HA1)<sup>39</sup> (Fig. 4d; Supplementary Movie 1).

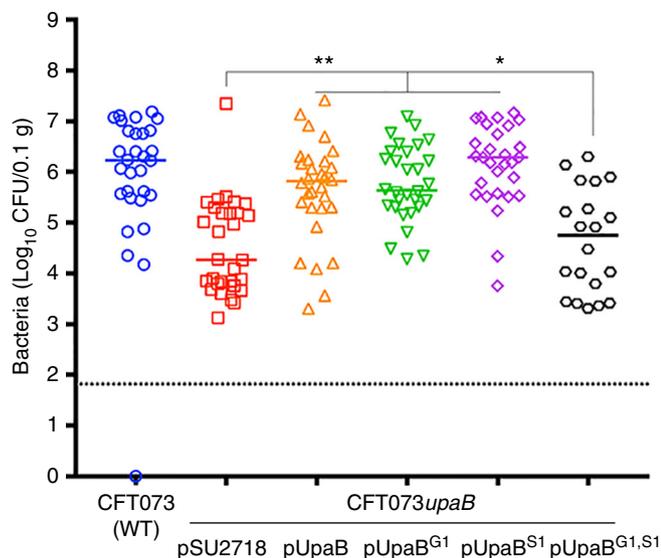


**Fig. 4** UpaB binds fibronectin type III. **a** Fibronectin domain organisation composed of 12 type I modules (FnI), 2 type II modules (FnII) and 15–17 type III modules (FnIII). Commercially available fragments used in this work include the heparin/gelatin FnI<sub>1–9</sub>, FnII<sub>1–2</sub>, the gelatin FnI<sub>6–9</sub>, FnII<sub>1–2</sub>, the cell binding FnIII<sub>2–11</sub> and the C-terminal heparin FnIII<sub>12–15</sub> fragments. **b** Binding of fibronectin fragments, as well as full-length (FL) fibronectin, to UpaB measured by enzyme-linked immunosorbent assay using an UpaB-specific polyclonal antibody. Data are shown as the mean  $\pm$  standard deviation of three replicates. **c** Tandem  $\beta$ -zipper interaction between Fn binding repeat peptides from *S. aureus* Fn-binding protein A (FnBPA) and Fn type I modules 2 and 3. The established mode of interaction between bacterial proteins and fibronectin involves the donation of up to 11 structurally disordered bacterial fibronectin repeats to form additional  $\beta$ -strands with consecutive FnI modules. **d** Model of the UpaB-FnIII interaction derived from NAMD simulations using the structures of UpaB and the FnIII<sub>1–2</sub> fragment (PDB: 2HA1), showing predominately hydrogen bonding between charged residues of UpaB (in particular, D246, D310, D336 and D375) and FnIII<sub>1</sub> (residues K32, R36, K40 and E70). The equivalent mutant simulation did not show any appreciable hydrogen bonding

To visualise this interaction, we also ran simulations with the  $\alpha^{UpaB\_S1}$  mutant that had lost its capacity to bind FN (Supplementary Movie 2). Modelling simulations were performed using NAMD 2.12<sup>43</sup> for a cumulative total of 1.2  $\mu$ s for each system (3 replicates of 400 ns each). Though the simulations are too short in timescale terms of protein–protein interactions to demonstrate specific binding, they do provide plausible binding mechanisms, showing that  $\alpha^{UpaB}$  could interact with FnIII via complementary charged residues without unfolding and/or donating  $\beta$ -strands. Specifically, our  $\alpha^{UpaB}$ -FnIII<sub>1–2</sub> simulations indicate that  $\alpha^{UpaB}$  primarily interacts with FnIII<sub>1</sub>, through hydrogen bond interactions mediated by  $\alpha^{UpaB}$  D246, R310, D336 and D375 residues with complementary charged areas on FnIII, particularly residues K32, K40, E70, R36 and of FnIII<sub>1</sub>. Substitutions of the  $\alpha^{UpaB}$  FnIII-interacting residues to alanine in the  $\alpha^{UpaB\_S1}$  mutant greatly reduced hydrogen bond interactions observed in the simulations (Supplementary Movie 2).  $\beta$ -Helix– $\beta$ -helix interactions have been observed previously for Ag43 and some trimeric ATs<sup>17,44,45</sup>. However, these studies show how an

AT  $\beta$ -helix can interact with another type of fold, namely the  $\beta$ -sandwich fold of FnIII.

**UpaB is highly conserved and immunogenic during UPEC infection.** To determine whether the structural features of UpaB required for binding FnIII and GAGs are conserved across the *E. coli* species, we screened the NCBI public database and an *in-house* collection, which were represented by 2818 draft and 199 complete *E. coli* genome sequences. Overall, the *upaB* gene was present in 1019 strains in this collection (34%) and was found in UPEC strains as well as intestinal pathogenic, commensal and other extra-intestinal pathogenic strains. Analysis of these 1019 translated UpaB amino acid sequences revealed that 95% (968/1019) shared an amino acid sequence identity >89%. Comparison of the translated UpaB amino acid sequence from seven completely sequenced UPEC strains showed that the putative GAG lyase active site was strictly conserved; there was also high conservation of the residues that contribute to the acidic groove, as well as the residues that interact with FnIII (Supplementary



**Fig. 5** UPEC colonisation of the mouse bladder is enhanced by UpaB GAG- and fibronectin-binding interactions. C57BL/6 mice were challenged transurethally with wild-type CFT073, CFT073 $upaB$ (pSU2718), CFT073 $upaB$ (pUpaB), CFT073 $upaB$ (pUpaB<sup>G1</sup>), CFT073 $upaB$ (pUpaB<sup>S1</sup>) and CFT073 $upaB$ (pUpaB<sup>G1,S1</sup>). The results represent log<sub>10</sub> CFU/0.1 g bladder tissue of individual mice at 24 h post-infection, and the horizontal bars mark group medians. A minimum of 20 mice were assessed per group (pooled from at least 2 independent experiments). Data were compared using Kruskal-Wallis analysis of variance (ANOVA) with Dunn's multiple comparisons correction (\* $P < 0.05$ ; \*\* $P < 0.01$ )

Fig. 8). We also examined whether an immunological response against UpaB was elicited during human UPEC infection and showed that plasma samples from urosepsis patients infected with UpaB-positive *E. coli* strains produced significantly higher anti- $\alpha^{UpaB}$  antibody titres compared to healthy individuals (Supplementary Fig. 9).

**Relevance of GAG and FN-binding regions in vivo.** We previously showed that mutation of *upaB* in CFT073 led to decreased bladder colonisation in experimental mice<sup>19</sup>. In order to examine how the GAG- and FN-binding properties of UpaB impact its function in vivo, we constructed plasmids containing the S1, G1 and double S1–G1 mutations in the full-length *upaB* gene. These plasmids were transformed into our *upaB* mutant strain (CFT073 $upaB$ ) to generate a set of strains with plasmid pSU2718 (vector control), pUpaB (wild-type (WT) UpaB), pUpaB<sup>G1</sup> (UpaB with mutated GAG-binding site), pUpaB<sup>S1</sup> (UpaB with mutated FN-binding site) or pUpaB<sup>G1,S1</sup> (UpaB with mutated GAG- and FN-binding sites). Next, we examined the capacity of our CFT073 $upaB$  complemented strains to colonise the mouse bladder. In these experiments, CFT073 $upaB$  complemented with pUpaB, pUpaB<sup>G1</sup> and pUpaB<sup>S1</sup> restored bladder colonisation at 24 h post-infection to a level equivalent to colonisation by WT CFT073 (Fig. 5). In contrast, complementation with either the vector control plasmid pSU2718 or pUpaB<sup>G1,S1</sup> did not restore bladder colonisation to WT levels, and these levels were significantly reduced at 24 h post-infection compared to colonisation by CFT073 $upaB$  containing pUpaB, pUpaB<sup>G1</sup> or pUpaB<sup>S1</sup> (Fig. 5). This lack of complementation by pUpaB<sup>G1,S1</sup> was not due to lack of expression of the mutant protein on the cell surface, as demonstrated by western blot analysis and whole-cell ELISA (Supplementary Fig. 10a, b). The stability of the pUpaB<sup>G1,S1</sup> mutant was also confirmed by purification and biophysical characterisation of recombinant  $\alpha^{UpaB\_G1,S1}$  (Supplementary

Fig. 3a and Supplementary Fig. 10c). A similar colonisation profile was observed for each of the UPEC strains in the urine of these experimentally infected mice (Supplementary Fig. 10d).

## Discussion

AT adhesins are a common group of proteins that play a central role in bacterial pathogenesis. They allow bacteria to adhere to human cells, aggregate with other bacteria and form biofilms, all key facilitators of bacterial virulence. Currently, our understanding of the function of ATs is limited, and detailed information at the level of atomic structure as well as the precise molecular mechanisms that govern their interaction with target molecules is lacking. Here we describe the structure and mechanism of action for UpaB, a recently described AT adhesin from UPEC.

UpaB differs from other characterised AIDA-I AT adhesins as it does not self-associate. Self-association of AIDA-I ATs on the bacterial cell surface is the mechanism by which AIDA-I ATs promote bacterial aggregation and biofilm formation<sup>17</sup>. The lack of this feature in UpaB is consistent with previous studies which showed that UpaB overexpression did not impact on these phenotypes<sup>19,46</sup>. Subsequent determination of the  $\alpha^{UpaB}$  structure revealed an unprecedented departure from the common  $\beta$ -helix fold previously determined for all other ATs and led to the definition of several novel features in the architecture of UpaB. Extensions of  $\beta$ -strands within  $\alpha^{UpaB}$  together with large loops create a long acidic groove that can bind GAG. This site bears structural similarity to characterised GAG lyases<sup>28</sup> and promoted the binding of UpaB to GAG-‘like’ substrates, thus supporting a potential polysaccharide-binding site in an AT protein.

We also show that UpaB can bind to human FN. Using a series of UpaB mutants in combination with different fragments of FN, we showed that this occurs via interaction with type III FN, most likely at FnIII<sub>1–2</sub>. Despite human FN being one of the most common targets for bacterial adhesins, a detailed mechanistic description of the structural basis for binding to FnIII is lacking<sup>32,38,40–42</sup>. In UpaB, this interaction involves the folded UpaB  $\beta$ -helix presenting a ladder of charged/polar residues that interact with complementary charges on FnIII. This is in contrast to previously characterised FnBP–FN interactions, which occur through the donation of a series of disordered repeats from the FnBP to form a tandem  $\beta$ -zipper interaction with consecutive FnI domains<sup>37,47</sup>. This mode of UpaB binding to FnIII might also be utilised by other bacterial proteins that bind FnIII. Indeed, binding interactions involving  $\beta$ -helices have been increasingly observed among ATs and other proteins<sup>17,44,45</sup>.

Our previous work showed that deletion of *upaB* reduces UPEC colonisation of the mouse bladder<sup>19</sup>. The in vivo data we present here show that mutating either the GAG-binding site or the FN-binding site does not affect UPEC colonisation, while mutating both binding sites leads to a modest decrease in colonisation. One possible interpretation of these results would be that UPEC colonisation may be facilitated by UpaB binding to both GAG and FN. However, it is possible that the multiple mutations required to inactivate both binding domains may have altered the stability or other functional features of the UpaB protein. Therefore, further work is needed to clarify the potential relevance of UpaB binding to GAG and FN during infection.

## Methods

**Bacteria and growth conditions.** *E. coli* strains and plasmids used in this study are listed in Supplementary Table 2. Bacteria were routinely grown at 37 °C on solid or in liquid Luria-Bertani (LB) medium supplemented ampicillin (100  $\mu$ g ml<sup>-1</sup>) or chloramphenicol (30  $\mu$ g ml<sup>-1</sup>).

**Cloning, expression and purification of  $\alpha$ UpaB.** The coding sequence for the *upaB* alpha domain (residues 38–500, locus tag c0426) was PCR-amplified from UPEC CFT073 genomic DNA using primers 4326-UpaB-F and 4327-UpaB-R containing ligation-independent cloning (LIC) overhangs (Supplementary Table 2). Using LIC cloning, the amplified gene was inserted into a modified version of a pMCSG7<sup>17</sup> vector, which encodes a N-terminal his<sub>6</sub>-tag followed by a thioredoxin (TRX) domain and a TEV protease cleavage site. The resulting plasmid, pUpaBa, was used for the expression of  $\alpha$ UpaB and introduces three residues at the N-terminus (i.e. SNA) upon removal of the his<sub>6</sub>-TRX-tag with TEV. The  $\alpha$ UpaB protein was expressed in *E. coli* BL21 (DE3) pLysS cells (Invitrogen) using autoinduction (24 h at 30 °C) in the presence of the appropriate antibiotics (ampicillin 100  $\mu$ g ml<sup>-1</sup>, chloramphenicol 34  $\mu$ g ml<sup>-1</sup>). Cells were harvested, resuspended in 25 mM Tris pH 7.5 and 150 mM NaCl and lysed by cell sonication. The lysate was cleared by centrifugation and loaded onto a HisTrap column (GE Healthcare). Proteins were eluted in a gradient of 0–500 mM imidazole. Fractions containing  $\alpha$ UpaB were cleaved with TEV protease and the uncleaved protein was removed by further nickel affinity chromatography. Size exclusion chromatography (Superdex S-75 GE Healthcare) in 25 mM Hepes and 150 mM NaCl pH 7.0, was used to further purify  $\alpha$ UpaB as assessed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

The pUpaB plasmid was used as the parent vector for the construction of all  $\alpha$ UpaB mutants, namely  $\alpha$ UpaB- $\Delta$ t6-10,  $\alpha$ UpaB- $\Delta$ t1-2,  $\alpha$ UpaB- $\Delta$ t3-4,  $\alpha$ UpaB- $\Delta$ t5-6,  $\alpha$ UpaB- $\Delta$ t7-8,  $\alpha$ UpaB-G1,  $\alpha$ UpaB-G2,  $\alpha$ UpaB-G3,  $\alpha$ UpaB-S1,  $\alpha$ UpaB-S2,  $\alpha$ UpaB-S3 and  $\alpha$ UpaB-G1,S1 (Supplementary Table 2). All constructs were generated by Epoch Life Science, confirmed by sequencing and transformed into *E. coli* BL21 DE3 pLysS (Supplementary Table 2). The  $\alpha$ UpaB mutants were expressed and purified as described for the native  $\alpha$ UpaB.

**Crystallisation.** Crystals of  $\alpha$ UpaB were grown at 20 °C using the hanging-drop vapour-diffusion technique. Crystals grew at 20 mg ml<sup>-1</sup> in 0.1 M sodium acetate pH 4.8, 0.2 M ammonium sulfate and 28% (w/v) PEG 4000. Crystals pre-equilibrated in reservoir solution containing 20% glycerol were flash-cooled in liquid nitrogen. Xenon derivatisation was performed using a Xenon chamber (Hampton Research) at 20 bar for 1 min before flash freezing.

**Structure determination and refinement.** Native data were collected ( $\lambda = 0.954$ , –163 °C) from a single crystal with an ADSC Q315r CCD detector on the MX2 micro-crystallography beamline at the Australian Synchrotron. The data were integrated and scaled with HKL2000<sup>48</sup>. Anomalous data were collected ( $\lambda = 1.3776$ , –163 °C) from 2 crystals at the MX2 beamline. This data was integrated, scaled and merged using XDS/XSCALE<sup>49</sup>. All crystals belonged to spacegroup P3<sub>1</sub>21 with similar cell dimensions of  $a \approx 69$  Å,  $b \approx 69$  Å,  $c \approx 166$  Å and  $\alpha = 90.0^\circ$ ,  $\beta = 90.0^\circ$  and  $\gamma = 120.0^\circ$ . This was consistent with one  $\alpha$ UpaB molecule per asymmetric unit. The structure of  $\alpha$ UpaB was determined by single isomorphous replacement using anomalous signal from Xenon. SHELX C,D,E<sup>50</sup> was used to find the Xenon atoms, phasing and density modification. Eight Xenon atoms were found per asymmetric unit. ARP/wARP<sup>51</sup> was used for initial model building against the experimental phases. This model underwent rounds of manual model building using the program COOT<sup>52</sup> and refinement using Refmac5<sup>53</sup> and phenix.refine<sup>54</sup> to 1.97 Å using native data. The quality of the model was monitored during refinement by the Rfree value, which represented 5% of the data. The structure was validated by the MolProbity<sup>55</sup> server and the figures were created with PyMOL<sup>56</sup>. Ramachandran statistics showed 97.87% of residues in the most favoured region and 2.13% in the allowed regions. Details of data-processing statistics and final refinement values are summarised in Table 1.

**Analytical ultracentrifugation.** Sedimentation velocity experiments were performed in a Beckman Coulter model XL-I analytical ultracentrifuge with a An50-Ti rotor. Double-sector quartz cells were loaded with 400  $\mu$ l of buffer (25 mM Tris pH 7.5 and 150 mM NaCl) and 380  $\mu$ l  $\alpha$ UpaB at 0.5, 1 and 2.2 mg ml<sup>-1</sup>. Initial scans were performed at 725  $\times g$  to determine the optimal wavelength and radial positions. Absorbance readings were collected at 280 nm and 128,794  $\times g$ . at 20 °C. Solvent density, solvent viscosity and estimates of the partial specific volume of  $\alpha$ UpaB (0.7203 ml g<sup>-1</sup>) at 20 °C were calculated with SEDNTERP<sup>57</sup>. Data were analysed using  $c(s)$  and  $c(M)$  with SEDFIT<sup>58</sup>.

**SAXS data collection and analysis.** Data were collected on the SAXS-WAXS beamline at the Australian Synchrotron. Serial dilutions of a 2.7 mg ml<sup>-1</sup> stock were made to give samples with concentrations between ~0.1 and 2.7 mg ml<sup>-1</sup>. All samples were centrifuged at 10,000  $\times g$  prior to being loaded into a 96-well plate. To minimise the effects of radiation damage, samples (~80  $\mu$ l) were maintained at 283 K and drawn into a capillary from the 96-well plate and flowed past the beam. All measured two-dimensional data were averaged and corrected for transmission, solvent scattering and detector sensitivity and radially averaged to produce  $I(q)$  vs.  $q$  profiles using Scatterbrain (v 2.7.1).

The estimated molecular masses were calculated using values for contrast and partial specific volume predicted from the protein sequence using MULCh (v 1.1)<sup>59</sup> along with the Porod volume. Data processing and Guinier analysis was performed using Primus (v 3.2)<sup>60</sup>. The pair-distance distribution function,  $p(r)$ , was generated

from the experimental data using GNOM (v 4.6)<sup>61</sup>, from which  $I(0)$ ,  $R_g$  and  $D_{max}$  were determined.

The program CORAL (v 1.1)<sup>62</sup> was used to generate 16 rigid-body models of the protein, where the missing N- and C-terminal residues from the crystal structure (PDB: 6BEA) were modelled as dummy residues. All models were qualitatively similar, and the model with the lowest  $\chi^2$  was chosen as the representative structure. The program DAMMIN (v 5.3)<sup>63</sup> was used to generate 16 molecular envelopes, which were averaged and filtered using the program DAMAVER (v 2.8)<sup>64</sup>. The SAXS data and models have been deposited in the SASBDB<sup>65</sup>.

**Polysaccharide lyase assay and gel.** Polysaccharide lyase assays were performed with human chondroitin sulfate A/C (Sigma), human chondroitin sulfate B (Sigma) or human heparin sulfate (Sigma) at 0.5 mg ml<sup>-1</sup> in 100 mM Tris 50 mM sodium acetate pH 8.0 and 10 mM CaCl<sub>2</sub> to final volume of 100  $\mu$ l. Purified  $\alpha$ UpaB was added to a final concentration of 0.05 mg ml<sup>-1</sup> along with the negative control Antigen 43, with the positive control chondroitin lyase ABC (Sigma) used at 0.005 mg ml<sup>-1</sup>. Assays were set-up in 96-well microplates (UV-Star Greiner) and substrate cleavage was followed by A<sub>232</sub> measurements every 4 min for 2 h at 37 °C, using an EnSpire 2300 multilabel reader (Perkin Elmer). Assay samples were then analysed by SDS-PAGE (4–12%) with Alician Blue/Silver staining.

**ELISA of glycosaminoglycans, FN, laminin or fibrinogen with  $\alpha$ UpaB.** ELISAs of the glycosaminoglycans human chondroitin sulfate A, B, C and heparin sulfate were performed by coating the molecules onto Nunc Maxisorp flat-bottom 96-well plates (Thermo Scientific) at 5, 10, 20 and 40  $\mu$ g ml<sup>-1</sup>. Plates were blocked with 1% w/v bovine serum albumin and probed with 10  $\mu$ g ml<sup>-1</sup> of  $\alpha$ UpaB. The binding of  $\alpha$ UpaB was detected using a UpaB-specific polyclonal antibody (1 in 500 dilution in phosphate-buffered saline (PBS))<sup>19</sup> followed by alkaline phosphate-conjugated goat anti-rabbit IgG (1 in 10,000 dilution in PBS (Sigma A3687)). For human laminin FN (10  $\mu$ g ml<sup>-1</sup>, Sigma), human fibrinogen (10  $\mu$ g ml<sup>-1</sup>, Sigma) and full-length human FN (10  $\mu$ g ml<sup>-1</sup>, Sigma), ELISAs were performed in the same manner. To detect binding to immobilised UpaB or UpaB mutants, ELISA plates were coated with purified  $\alpha$ UpaB, truncates or  $\alpha$ UpaB surface/groove mutants (10  $\mu$ g ml<sup>-1</sup>) and probed with 10  $\mu$ g ml<sup>-1</sup> each of full-length human FN or FN fragments; including the gelatin-binding (Sigma), heparin/gelatin (Sigma), cell-binding (Merck Millipore) or C-terminal heparin-binding (Merck Millipore)  $\alpha$ -chymotryptic FN fragments. The binding of FN and FN fragments was detected using anti-FN antibody (1 in 1000 dilution in PBS (Sigma F3648)) followed by alkaline phosphate-conjugated goat anti-rabbit IgG (1 in 10,000 dilution in PBS (Sigma A3687)). The reaction was developed in the presence of alkaline phosphatase substrate (Sigma) and absorbance was read at 405 nm.

**Fluorescence thermal shift assays.** Fluorescence thermal shift assay was conducted in 384-well plate format with an assay volume of 10  $\mu$ l. Recombinant  $\alpha$ UpaB protein sample and 5000 $\times$  Sypro-Orange (Invitrogen) were diluted and mixed in a Hepes buffer (20 mM HEPES, 150 mM NaCl, pH 7.5) to a protein concentration of 0.5  $\mu$ g per well and 5 $\times$  Sypro-Orange. Because of the low signal intensity of  $\alpha$ UpaB, higher than usual protein concentration was used. After adding the protein Sypro-Orange mix to a PCR plate, testing compounds were added using a Labcyte Echo550 acoustic liquid transfer robot. Plates were mixed, sealed with optical clear plastic seal and centrifuged. Thermal scanning coupled with fluorescence detection was performed on a CFX384 qPCR machine at 1.5 °C min<sup>-1</sup> from 10 °C to 85 °C. Data analysis was performed using the in-house software excelFITS, which uses IDBS XLfit for fitting the fluorescence data to a Boltzmann function to determine the melting temperature  $T_m$  and other thermal transition parameters. Two compound libraries of 2700 molecules and 88 molecules each were screened with  $\alpha$ UpaB. No hit compounds were found from the 2700 molecule Spectrum collection of known and experimental drugs and natural products (MicroSource Discovery). The screen of 88 carbohydrates yielded two hits.

**Circular dichroism spectroscopy.** An Aviv model 420 Circular Dichroism Spectrometer was used to investigate the structural properties of  $\alpha$ UpaB proteins at 0.3 mg ml<sup>-1</sup> in 25 mM Hepes pH 7.0 and 150 mM NaCl. Wavelength scans were performed with 0.5 nm steps between 200 to 250 nm at 20 °C.

**ELISA of whole cells expressing UpaB with FN.** Full-length *upaB* from UPEC CFT073 was cloned into plasmid pSU2718 at *Xba*I-*Hind*III restriction sites with primers 6460-UpaBF1 and 6461-UpaBR (Supplementary Table 2). The resultant parent plasmid (pUpaB) was used for construction of all UpaB deletion mutants used in this study (Supplementary Table 2). Specific mutations and deletions were introduced into *upaB* by Epoch Life Science to generate the following plasmids [pUpaB<sup>G1</sup>, pUpaB<sup>S1</sup>, pUpaB <sup>$\Delta$ t1-2</sup>, pUpaB <sup>$\Delta$ t3-4</sup>, pUpaB <sup>$\Delta$ t5-6</sup> and pUpaB <sup>$\Delta$ t7-8</sup>]. All constructs were confirmed by sequencing and transformed into MS427 (Supplementary Table 2). For all assays, overnight cultures of bacterial cells were normalised to an optical density at 600 nm (OD<sub>600 nm</sub>) of 1.0. Whole-cell ELISAs were performed as described for purified UpaB proteins. Whole-cell ELISAs for the detection of UpaB and UpaB deletion mutants on the *E. coli* cell surface were performed using polyclonal anti- $\alpha$ UpaB antibody (1 in 500 dilution in PBS) and detected using alkaline phosphate-conjugated goat anti-rabbit IgG (1 in 10,000

dilution in PBS (Sigma A3687)). The interaction of UpaB deletion mutants with FN was examined by whole-cell ELISA using anti-FN antibody (1 in 1000 dilution in PBS (Sigma F3648)) and detected using alkaline phosphate-conjugated goat anti-rabbit IgG (1 in 10,000 dilution in PBS (Sigma A3687)).

**SPR measurements.** A Biacore T200 biosensor instrument was used to measure the affinity of the interaction of UpaB with full-length human FN. FN was covalently immobilised onto a CM5 chip at two different densities, 1000 RU and 5000 RU, using amine coupling method. SPR experiments were performed at 25 °C using PBS-T (1× PBS pH 7.4 and 0.05% Tween 20) as the running buffer. To generate binding data, UpaB at concentrations ranging from 100 to 0.8 μM was injected over immobilised FN at a constant flow rate of 90 ml min<sup>-1</sup> for 50 s; UpaB dissociation was monitored by flowing running buffer at 90 ml min<sup>-1</sup> for 150 s. The surface was regenerated after each cycle by injecting 0.1% SDS. Steady-state equilibrium analysis was carried out using the Biacore T200 evaluation software.  $K_D$  is expressed as mean ± standard error of the mean (SEM). Experiments were conducted on three independent occasions with fresh immobilisation.

**Mouse infections.** Plasmids containing the full-length *upaB* gene with mutations in the GAG-binding site (pUpaB<sup>G1</sup>), FN-binding site (pUpaB<sup>S1</sup>) and both binding sites (pUpaB<sup>G1, S1</sup>) were generated by Epoch Life Science. These plasmids, together with a plasmid containing the full-length *upaB* gene (pUpaB) and the vector control (pSU2718), were transformed into CFT073*upaB*, respectively, to generate the strains used in the mouse UTI model. Female C57BL/6 mice aged 10–12 weeks (Animal Resources Centre) were inoculated by transurethral infection to deliver approximately 10<sup>8</sup> bacteria to the bladder, as described elsewhere<sup>66</sup>. Briefly, mice ( $n = 10$  per group) were anaesthetised by inhalation exposure to isoflurane and a sterile Teflon-coated catheter attached to a 1 ml syringe was used to deliver 50 μl of PBS containing ~10<sup>8</sup> colony-forming units (c.f.u.) of bacteria (at a rate of 5 μl s<sup>-1</sup>) to the bladder. Urine samples were collected 24 h after challenge and were diluted in PBS and plated on LB agar or LB agar supplemented with 30 μg ml<sup>-1</sup> chloramphenicol, as appropriate, for colony counts. Subsequently, the mice were euthanised (using isoflurane overdose followed by cervical dislocation), and the bladders and kidneys were collected, weighed and homogenised in sterile PBS. The tissue homogenates were diluted and plated on agar as above to quantify c.f.u. per 0.1 g tissue. Data are compiled from at least two independent experiments.

**ELISA to detect anti-αUpaB antibodies from urosepsis patients.** Blood plasma was collected from 45 patients presenting with urosepsis at the Princess Alexandra Hospital (Brisbane, Australia) along with matching blood culture UPEC isolates. Isolates were screened for the presence of *upaB*, resulting in 33 plasma samples with matching infecting strains positive for *upaB* to be assessed. Forty-two plasma samples from healthy volunteers were obtained as controls. Recombinant UpaB α-domain (αUpaB; 10 μg ml<sup>-1</sup>) was coated onto Nunc Maxisorp flat-bottom 96-well plates (Thermo Scientific), plasma samples were added and peroxidase-conjugated anti-human IgG (1:30,000 dilution in 5% skim milk (Sigma A0170)) was applied as a secondary antibody for detection (incubated at 37 °C for 90 min). Plates were developed with 3,3',5,5'-tetramethylbenzidine with absorbance determined using a SpectraMax 190 Absorbance Microplate Reader at 450 nm. Statistical analysis between patient and healthy plasma was performed using an unpaired two-sample *t* test.

**Sequence analysis.** The *upaB* sequences from the NCBI database and *in house* collection were compared and aligned using CLC Main Workbench (Qiagen).

**Modelling.** Models for dynamics simulations were constructed for both αUpaB and αUpaB<sub>S1</sub> using the UpaB crystal structure and the FnIII<sub>1-2</sub> crystal structure (2HA1). UpaB models were positioned approximately 10 Å from the FN fragment and solvated with TIP3P and electrically neutralised with 0.15 M sodium chloride. The models initial dimensions of 108 × 81 × 120 Å, contained 100,010 and 100,015, atoms respectively. Each model was run independently 3 times for 400 ns, performed with NAMD2.12.34, for a cumulative total of 2.4 μs of simulation. Simulations were performed with NAMD2.12.43. Long-range Coulomb forces were computed with the Particle Mesh Ewald method with a grid spacing of 1 Å. Two-fs timesteps were used with non-bonded interactions calculated every 2 fs and full electrostatics every 4 fs while hydrogens were constrained with the SHAKE algorithm. The cut-off distance was 12 Å with a switching distance of 10 Å and a pair-list distance of 14 Å. The temperature was set to 310 K. Pressure was controlled to 1 atmosphere using the Nosé–Hoover Langevin piston method employing a piston period of 100 fs and a piston decay of 50 fs. Trajectory frames were captured every 100 ps. Simulation trajectories were viewed and mapped with VMD<sup>67</sup>.

Docking of the GAG into UpaB utilised a model based on the NMR structure of unsulfated chondroitin (PDB: 2KQO). Docking of the GAG model was performed against the UpaB structure using Autodock Vina<sup>68</sup>. A 52 × 24 × 46 Å search space was set up to cover the UpaB surface and groove. Standard chemical bond torsions were applied to the GAG molecule and UpaB was kept rigid apart from the following residues: N189, Q197, K256, and K343. Docking conformations were ranked against the predicted free energy of binding (kcal mol<sup>-1</sup>).

**Ethics statement.** All animal experimentation was conducted in accordance with the guidelines of the National Health and Medical Research Council. The Griffith University Animal Ethics Committee approved this study (MSC/01/18/AEC). The use of human blood plasma from patients was approved by the institutional review board of the Princess Alexandra Hospital (2008/264). The need for patient informed consent was waived, as the primary purpose for the collection of these samples was for other diagnostic procedures, and all patient information was de-identified. The collection of human blood from control subjects was approved by the institutional review board of Griffith University (MSC/18/10/HREC). Informed consent was obtained from all control subjects.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Coordinates and structure-factor files have been deposited in the Protein Data Bank, with accession code 6BEA. Scattering data and models have been deposited in the SASDBB with accession code SASDC45.

Received: 13 February 2019 Accepted: 28 March 2019

Published online: 29 April 2019

## References

- Kline, K. A., Falker, S., Dahlberg, S., Normark, S. & Henriques-Normark, B. Bacterial adhesins in host-microbe interactions. *Cell Host Microbe* **5**, 580–592 (2009).
- Ribet, D. & Cossart, P. How bacterial pathogens colonize their hosts and invade deeper tissues. *Microbes Infect.* **17**, 173–183 (2015).
- Vo, J. L. et al. Autotransporter adhesins in *Escherichia coli* pathogenesis. *Proteomics* <https://doi.org/10.1002/pmic.201600431> (2017).
- Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C. & Ala'Aldeen, D. Type V protein secretion pathway: the autotransporter story. *Microbiol. Mol. Biol. Rev.* **68**, 692–744 (2004).
- Benz, I. & Schmidt, M. A. Structures and functions of autotransporter proteins in microbial pathogens. *Int. J. Med. Microbiol.* **301**, 461–468 (2011).
- Hartmann, M. D. et al. Complete fiber structures of complex trimeric autotransporter adhesins conserved in enterobacteria. *Proc. Natl Acad. Sci. USA* **109**, 20907–20912 (2012).
- Linke, D., Riess, T., Autenrieth, I. B., Lupas, A. & Kempf, V. A. Trimeric autotransporter adhesins: variable structure, common function. *Trends Microbiol.* **14**, 264–270 (2006).
- Cotter, S. E., Surana, N. K. & St Geme, J. W. III. Trimeric autotransporters: a distinct subfamily of autotransporter proteins. *Trends Microbiol.* **13**, 199–205 (2005).
- Emsley, P., Charles, I. G., Fairweather, N. F. & Isaacs, N. W. Structure of *Bordetella pertussis* virulence factor P.69 pertactin. *Nature* **381**, 90–92 (1996).
- Otto, B. R. et al. Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic *Escherichia coli*. *J. Biol. Chem.* **280**, 17339–17345 (2005).
- Khan, S., Mian, H. S., Sandercock, L. E., Chirgadze, N. Y. & Pai, E. F. Crystal structure of the passenger domain of the *Escherichia coli* autotransporter EspP. *J. Mol. Biol.* **413**, 985–1000 (2011).
- Meng, G., Spahich, N., Kenjale, R., Waksman, G. & St Geme, J. W. III. Crystal structure of the Haemophilus influenzae Hap adhesin reveals an intercellular oligomerization mechanism for bacterial aggregation. *EMBO J.* **30**, 3864–3874 (2011).
- Wells, T. J., Totsika, M. & Schembri, M. A. Autotransporters of *Escherichia coli*: a sequence-based characterization. *Microbiology* **156**, 2459–2469 (2010).
- Henderson, I. R., Meehan, M. & Owen, P. Antigen 43, a phase-variable bipartite outer membrane protein, determines colony morphology and autoaggregation in *Escherichia coli* K-12. *FEMS Microbiol. Lett.* **149**, 115–120 (1997).
- Danese, P. N., Pratt, L. A., Dove, S. L. & Kolter, R. The outer membrane protein, antigen 43, mediates cell-to-cell interactions within *Escherichia coli* biofilms. *Mol. Microbiol.* **37**, 424–432 (2000).
- Klemm, P., Hjerrild, L., Gjermansen, M. & Schembri, M. A. Structure-function analysis of the self-recognizing Antigen 43 autotransporter protein from *Escherichia coli*. *Mol. Microbiol.* **51**, 283–296 (2004).
- Heras, B. et al. The antigen 43 structure reveals a molecular Velcro-like mechanism of autotransporter-mediated bacterial clumping. *Proc. Natl Acad. Sci. USA* **111**, 457–462 (2014).

18. Foxman, B. Urinary tract infection syndromes: occurrence, recurrence, bacteriology, risk factors, and disease burden. *Infect. Dis. Clin. North Am.* **28**, 1–13 (2014).
19. Allsopp, L. P. et al. Molecular characterization of UpaB and UpaC, two new autotransporter proteins of uropathogenic *Escherichia coli* CFT073. *Infect. Immun.* **80**, 321–332 (2012).
20. Klemm, P., Vejborg, R. M. & Sherlock, O. Self-associating autotransporters, SAATs: functional and structural similarities. *Int. J. Med. Microbiol.* **296**, 187–195 (2006).
21. Zhu-Ge, X. K. et al. The effects of upaB deletion and the double/triple deletion of upaB, aatA, and aatB genes on pathogenicity of avian pathogenic *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **99**, 10639–10654 (2015).
22. Oliver, D. C., Huang, G., Nodel, E., Pleasance, S. & Fernandez, R. C. A conserved region within the *Bordetella pertussis* autotransporter BrkA is necessary for folding of its passenger domain. *Mol. Microbiol.* **47**, 1367–1383 (2003).
23. Kajava, A. V. & Steven, A. C. The turn of the screw: variations of the abundant beta-solenoid motif in passenger domains of Type V secretory proteins. *J. Struct. Biol.* **155**, 306–315 (2006).
24. Xu, F. et al. Novel molecular insights into the catalytic mechanism of marine bacterial alginate lyase AlyGC from polysaccharide lyase family 6. *J. Biol. Chem.* **292**, 4457–4468 (2017).
25. Pickersgill, R., Smith, D., Worboys, K. & Jenkins, J. Crystal structure of polygalacturonase from *Erwinia carotovora* ssp. *carotovora*. *J. Biol. Chem.* **273**, 24660–24664 (1998).
26. Rozeboom, H. J., Beldman, G., Schols, H. A. & Dijkstra, B. W. Crystal structure of endo-xylogalacturonan hydrolase from *Aspergillus tubingensis*. *FEBS J.* **280**, 6061–6069 (2013).
27. Michel, G. et al. The structure of chondroitin B lyase complexed with glycosaminoglycan oligosaccharides unravels a calcium-dependent catalytic machinery. *J. Biol. Chem.* **279**, 32882–32896 (2004).
28. Garron, M. L. & Cygler, M. Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology* **20**, 1547–1573 (2010).
29. Hurst, R. E. Structure, function, and pathology of proteoglycans and glycosaminoglycans in the urinary tract. *World J. Urol.* **12**, 3–10 (1994).
30. Hook, M., Kjellen, L. & Johansson, S. Cell-surface glycosaminoglycans. *Annu. Rev. Biochem.* **53**, 847–869 (1984).
31. Walia, B. et al. Polarized fibronectin secretion induced by adenosine regulates bacterial-epithelial interaction in human intestinal epithelial cells. *Biochem. J.* **382**, 589–596 (2004).
32. Mullen, L. M. et al. Novel adhesin from *Pasteurella multocida* that binds to the integrin-binding fibronectin FnIII9-10 repeats. *Infect. Immun.* **76**, 1093–1104 (2008).
33. Raibaud, S. et al. *Borrelia burgdorferi* binds fibronectin through a tandem beta-zipper, a common mechanism of fibronectin binding in staphylococci, streptococci, and spirochetes. *J. Biol. Chem.* **280**, 18803–18809 (2005).
34. Casillas-Ituarte, N. N. et al. Amino acid polymorphisms in the fibronectin-binding repeats of fibronectin-binding protein A affect bond strength and fibronectin conformation. *J. Biol. Chem.* **292**, 8797–8810 (2017).
35. Lin, Y. P. et al. A novel fibronectin type III module binding motif identified on C-terminus of *Leptospira immunoglobulin-like protein*, LigB. *Biochem. Biophys. Res. Commun.* **389**, 57–62 (2009).
36. Cole, D. K. et al. Human TCR-binding affinity is governed by MHC class restriction. *J. Immunol.* **178**, 5727–5734 (2007).
37. Schwarz-Linek, U. et al. Pathogenic bacteria attach to human fibronectin through a tandem beta-zipper. *Nature* **423**, 177–181 (2003).
38. Henderson, B., Nair, S., Pallas, J. & Williams, M. A. Fibronectin: a multidomain host adhesin targeted by bacterial fibronectin-binding proteins. *FEMS Microbiol. Rev.* **35**, 147–200 (2011).
39. Vakonakis, I., Staunton, D., Rooney, L. M. & Campbell, I. D. Interdomain association in fibronectin: insight into cryptic sites and fibrillogenesis. *EMBO J.* **26**, 2575–2583 (2007).
40. Spahich, N. A. et al. Structural determinants of the interaction between the *Haemophilus influenzae* Hap autotransporter and fibronectin. *Microbiology* **160**, 1182–1190 (2014).
41. Christner, M. et al. The giant extracellular matrix-binding protein of *Staphylococcus epidermidis* mediates biofilm accumulation and attachment to fibronectin. *Mol. Microbiol.* **75**, 187–207 (2010).
42. Kingsley, R. A. et al. Fibronectin binding to the *Salmonella enterica* serotype Typhimurium ShdA autotransporter protein is inhibited by a monoclonal antibody recognizing the A3 repeat. *J. Bacteriol.* **186**, 4931–4939 (2004).
43. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
44. Nummelin, H. et al. The *Yersinia* adhesin YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. *EMBO J.* **23**, 701–711 (2004).
45. Govaerts, C., Wille, H., Prusiner, S. B. & Cohen, F. E. Evidence for assembly of prions with left-handed beta-helices into trimers. *Proc. Natl Acad. Sci. USA* **101**, 8342–8347 (2004).
46. Zude, I., Leimbach, A. & Dobrindt, U. Prevalence of autotransporters in *Escherichia coli*: what is the impact of phylogeny and pathotype? *Int. J. Med. Microbiol.* **304**, 243–256 (2014).
47. Bingham, R. J. et al. Crystal structures of fibronectin-binding sites from *Staphylococcus aureus* FnBPA in complex with fibronectin domains. *Proc. Natl Acad. Sci. USA* **105**, 12254–12258 (2008).
48. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
49. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
50. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485 (2010).
51. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 (2008).
52. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
53. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
54. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948–1954 (2002).
55. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
56. DeLano, W. L. *The PyMOL Molecular Graphics System*, <http://www.pymol.org>. (DeLano Scientific, San Carlos, CA, 2002).
57. Lebowitz, J., Lewis, M. S. & Schuck, P. Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci.* **11**, 2067–2079 (2002).
58. Schuck, P. et al. SEDFIT-MSTAR: molecular weight and molecular weight distribution analysis of polymers by sedimentation equilibrium in the ultracentrifuge. *Analyst* **139**, 79–92 (2014).
59. Whitten, A. E., Cai, S. & Trehwella, J. MULCh: modules for the analysis of small-angle neutron contrast variation data from biomolecular assemblies. *J. Appl. Crystallogr.* **41**, 222–226 (2008).
60. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
61. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
62. Petoukhov, M. V. et al. New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 (2012).
63. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).
64. Volkov, V. V. & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Cryst.* **36**, 860–864 (2003).
65. Valentini, E., Kikhney, A. G., Previtali, G., Jeffries, C. M. & Svergun, D. I. SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Res.* **43**, D357–D363 (2015).
66. Ulett, G. C. et al. Functional analysis of antigen 43 in uropathogenic *Escherichia coli* reveals a role in long-term persistence in the urinary tract. *Infect. Immun.* **75**, 3233–3244 (2007).
67. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph* **14**, 27–28 (1996).
68. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).

## Acknowledgements

This work was supported by an Australian Research Council (ARC) project grant (DP150102287, DP180102987), a La Trobe University Research Focus Area Understanding Disease Express Grant and a Victorian Life Sciences Computation Initiative grant (LTU0011) on its Peak Computing Facility at the University of Melbourne, an initiative of the Victorian Government, Australia. B.H. is supported by an ARC Future Fellowship (FT130100580) and M.A.S. by an Australian National Health and Medical Research Council Senior Research Fellowship (GNT1106930). We thank Dr. Salvatore Nocadello for getting us in contact with Professor Chi-Hao Luan. We thank Lahiru Katupitiya and Dean Gosling for excellent technical assistance. We acknowledge use of the MX2 and SAXS beamlines at the Australian Synchrotron and the CSIRO Collaborative Crystallisation Centre ([www.csiro.au/C3](http://www.csiro.au/C3); Melbourne, Australia).

## Author contributions

B.H., M.A.S. and J.J.P. designed the research. J.J.P. and A.W.L. performed the majority of the research. B.H., J.J.P., A.W.L., M.A.S., M.J.S. and G.C.U. analysed the data and wrote

the paper. Additionally, S.P. and C.L.G. contributed to X-ray data phasing; M.K. carried out molecular dynamics; A.E.W. processed and analysed SAXS data; C.-H.L. performed ligand screening; D.G.M. performed the bioinformatics analysis; G.W. performed the SPR analysis; K.M.P. and M.-D.P. performed expression/binding assays and generated full-length *upaB* constructs; L.T. performed immunogenicity experiments; M.J.S. and G.C.U. performed mouse experiments.

### Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-019-09814-6>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Journal peer review information:** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019