

# SCIENTIFIC REPORTS



OPEN

## Glyco-engineered cell line and computational docking studies reveals enterotoxigenic *Escherichia coli* CFA/I fimbriae bind to Lewis a glycans

Lynda Mottram<sup>1</sup>, Jining Liu<sup>2</sup>, Sonali Chavan<sup>3</sup>, Joshua Tobias<sup>1,4</sup>, Ann-Mari Svennerholm<sup>1</sup> & Jan Holgersson<sup>2</sup>

We have previously reported clinical data to suggest that colonization factor I (CFA/I) fimbriae of enterotoxigenic *Escherichia coli* (ETEC) can bind to Lewis a (Le<sup>a</sup>), a glycan epitope ubiquitous in the small intestinal mucosa of young children (<2 years of age), and individuals with a genetic mutation of *FUT2*. To further elucidate the physiological binding properties of this interaction, we engineered Chinese Hamster Ovary (CHO-K1) cells to express Le<sup>a</sup> or Le<sup>b</sup> determinants on both N- and O-glycans. We used our glyco-engineered CHO-K1 cell lines to demonstrate that CfaB, the major subunit of ETEC CFA/I fimbriae, as well as four related ETEC fimbriae, bind more to our CHO-K1 cell-line expressing Le<sup>a</sup>, compared to cells carrying Le<sup>b</sup> or the CHO-K1 wild-type glycan phenotype. Furthermore, using *in-silico* docking analysis, we predict up to three amino acids (Glu<sup>25</sup>, Asn<sup>27</sup>, Thr<sup>29</sup>) found in the immunoglobulin (Ig)-like groove region of CfaB of CFA/I and related fimbriae, could be important for the preferential and higher affinity binding of CFA/I fimbriae to the potentially structurally flexible Le<sup>a</sup> glycan. These findings may lead to a better molecular understanding of ETEC pathogenesis, aiding in the development of vaccines and/or anti-infection therapeutics.

Enterotoxigenic *Escherichia coli* (ETEC) is a leading cause of severe diarrhoeal illness in young children (<5 years of age) in low and middle-income countries. It is also a major cause of traveller's diarrhoea to ETEC endemic areas<sup>1,2</sup>. The bacterium has evolved to produce one or more of at least 23 distinct fimbrial (known as colonisation factors, CFs) or non-fimbrial adhesins, enabling ETEC to bind to the small intestinal epithelium before producing diarrhoeagenic enterotoxin(s)<sup>3</sup>. Thus, ETEC adherence factors are prerequisites for the initiation of pathogenesis, representing a critical point at which ETEC infections could be prevented<sup>4</sup>.

In previous clinical studies, we have demonstrated that Bangladeshi children expressing the histo-blood group antigen (HBGA) Lewis a (Le<sup>a</sup>, Le(a+b-)) phenotype, *FUT2* non-secretor status) are more likely to have symptomatic ETEC infection compared to children expressing the HBGA Lewis b (Le<sup>b</sup>, Le(a-b+)) phenotype, *FUT2* secretor status<sup>5,6</sup>. Interestingly, we have also observed that Bangladeshi children with the Le(a+b-) phenotype are more likely to be infected by ETEC expressing the colonisation factor antigen I (CFA/I) and the related ETEC CF family fimbriae or pili<sup>6</sup>. The likely explanation for this being, CFA/I could bind to Le<sup>a</sup> glycolipid structures present in the small intestinal mucosal layer of very young children (<2 years of age) and individuals with *FUT2* non-secretor status<sup>7,8</sup>.

CFA/I was the first human specific immunogenic ETEC CF to be described. It is a representative member of the antigenically defined ETEC CF class 5 pili, which are also commonly referred to as the α clade fimbrial usher

<sup>1</sup>Department of Microbiology and Immunology at the Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. <sup>2</sup>Department of Clinical Chemistry and Transfusion Medicine at the Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. <sup>3</sup>Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg, Sweden. <sup>4</sup>Institute of Specific Prophylaxis and Tropical Medicine, Medical University of Vienna, Vienna, Austria. Lynda Mottram and Jining Liu contributed equally to this work. Correspondence and requests for materials should be addressed to L.M. (email: [lynda.mottram@gu.se](mailto:lynda.mottram@gu.se))

protein (FUP) family<sup>4,9</sup>. Together, this ETEC CF group (CFA/I, CS1, CS2, CS4, CS14, CS17, CS19 and PCF071) accounts for the largest group of human specific ETEC CF expressing strains causing diarrhoeal disease worldwide<sup>2,4</sup>. Like other ETEC CF family members, CFA/I is comprised of a four gene operon, encoding for a long rigid homopolymorphic shaft with >1,000 copies of a major subunit (CfaB), with one or a few copies of the tip residing minor subunit (CfaE)<sup>4</sup>.

With regard to ETEC CFA/I binding to host cells, the minor subunit CfaE binds to the surface of erythrocytes<sup>4,10</sup>. The major subunit CfaB has been shown to bind to glycosphingolipids and human small intestinal glycolipid structures, such as those expressing Le<sup>a</sup> or asialo-GM1<sup>7</sup>. It has also been shown that specific monoclonal antibodies raised against CfaB inhibits ETEC CFA/I binding to cultured intestinal epithelial cells<sup>11–13</sup>. Moreover, an antibody that reacts strongly with the first 25 amino acids of the N-terminal fragment of CfaB has been shown to inhibit ETEC CFA/I bacterial adhesion to human jejunal enterocytes<sup>14,15</sup>. In contrary, it has been reported by others that CfaE of ETEC CFA/I binds to intestinal tissue and asialo-GM1 glycans that are expressed on erythrocytes and cultured intestinal epithelial cells<sup>10,16</sup>.

X-ray structural analysis has revealed CfaE and CfaB to have similar barrel like structures, with CfaE containing two and CfaB possessing one exposed hydrophobic immunoglobulin (Ig)-like fold(s), that structurally interact and complement each other<sup>9</sup>. Interestingly, a 12-amino acid stretch of the CfaB Ig-like fold (V<sup>24</sup>EKNITVTASVD<sup>35</sup>) that is located in the N terminal fragment of CfaB, shares structural similarities with all ETEC CF major subunits of the type 5 pili family. This 12 amino acid stretch of CfaB also shared structural similarities with class 1 pili from bacteria that can cause urinary and respiratory infections by binding to host glycolipids containing HBGAs<sup>9,17</sup>.

The aim of the present study was to create glycan defined Chinese hamster ovary (CHO-K1) cell line models of the human small intestinal mucosa, and to study the binding capabilities of ETEC CFA/I and the related CFs to Lewis Le<sup>a</sup> and Le<sup>b</sup> antigens expressed on the cell surface. We also perform computational molecular docking analysis to help understand why CFA/I binds to Le<sup>a</sup> but not Le<sup>b</sup> expressing glycans, as well as potentially identify novel CFA/I Le<sup>a</sup> glycan binding sites.

## Results

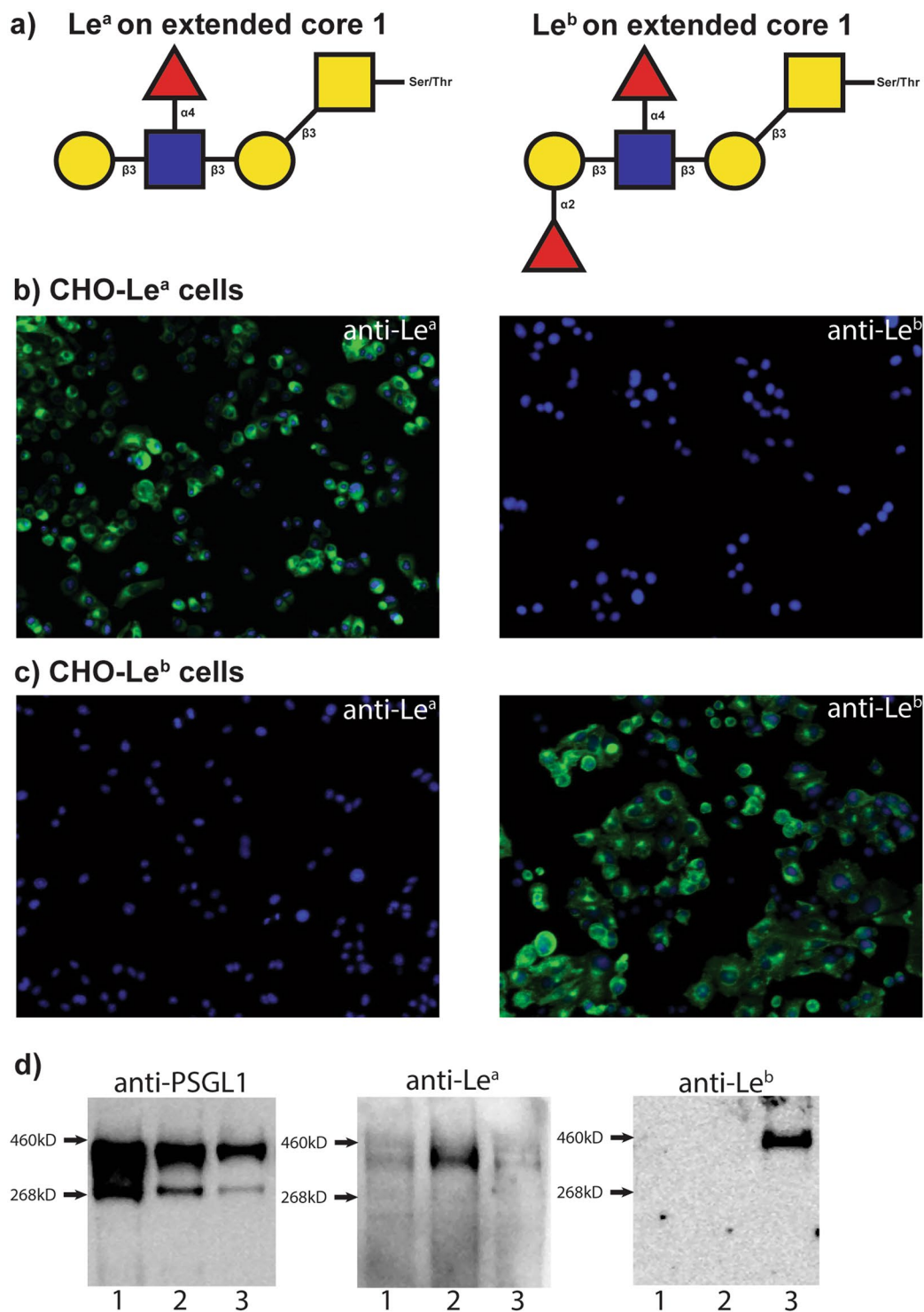
**Glyco-engineered CHO-K1 cells were produced expressing either Le<sup>a</sup> or Le<sup>b</sup>.** ETEC colonises the epithelial surface of the small intestinal mucosa, where intestinal villi and crypts express abundant Le<sup>a</sup> and/or Le<sup>b</sup> glycans<sup>4,18,19</sup>. To create defined HBGA Le<sup>a</sup> and Le<sup>b</sup> glycan models of the human small intestinal mucosa, CHO-K1 cells expressing the P-selectin glycoprotein ligand-1/immunoglobulin fusion protein (PSGL-1/mIgG2b; CHO-CP55 cells), were co-transfected with plasmids encoding: the extended core 1 glycan (GlcNAc $\beta$ 3Gal $\beta$ 3GalNAc $\alpha$ ) enzyme B3GNT3, the type 1 chain glycan (Gal $\beta$ 3GlcNAc) encoding enzyme B3GALT5, and the Lewis gene-encoding enzyme FUT3 alone (generating Le<sup>a</sup> cells; CHO-PSGL-Le<sup>a</sup>-a1) or together with the H gene-encoding FUT1 (generating Le<sup>b</sup>; CHO-PSGL-Le<sup>b</sup>-b1) (see Supplementary Figure S1 and Table S1 online, and Materials and Methods for further details)<sup>20–24</sup>.

Immunocytochemistry with anti-Le<sup>a</sup> and anti-Le<sup>b</sup> antibodies showed that all DAPI stained cells of the selected and expanded clone CHO-PSGL1-Le<sup>a</sup>-a1 were positively stained (in green) for Le<sup>a</sup>, but not for Le<sup>b</sup> (Fig. 1b). Similarly, the DAPI stained CHO-PSGL1-Le<sup>b</sup>-b4 clone was only positively stained for Le<sup>b</sup> (Fig. 1c). Subsequent analysis by SDS-PAGE and Western blotting of the purified PSGL-1/mIgG2b produced by these clones, revealed CHO-PSGL1-Le<sup>a</sup>-a1 expressed only Le<sup>a</sup>, whereas the CHO-PSGL1-Le<sup>b</sup>-b4 clone predominantly expressed Le<sup>b</sup> (Fig. 1d and Supplementary Fig S7 online). Therefore, the CHO-PSGL1-Le<sup>a</sup>-a1 (CHO-Le<sup>a</sup>), CHO-PSGL1-Le<sup>b</sup>-b4 (CHO-Le<sup>b</sup>) cell lines along with the Lewis antigen negative control cell lines CHO-CP55 and CHO-K1, were used for the ETEC CF-binding characterization experiments.

**ETEC CFA/I fimbriae binds to CHO-Le<sup>a</sup> cells.** In previous studies, we have demonstrated that ETEC CFA/I fimbriae bind to Le<sup>a</sup>-5 glycolipids separated onto thin layer chromatograms<sup>7</sup>. To evaluate the binding specificity of ETEC CFA/I fimbriae to defined Le<sup>a</sup> or Le<sup>b</sup> expressing glycans, the glyco-engineered CHO-K1 cells expressing either Le<sup>a</sup> (CHO-Le<sup>a</sup>) or Le<sup>b</sup> (CHO-Le<sup>b</sup>) determinants on their cell surface, along with the negative glycan phenotype control cell lines (CHO-CP55 and CHO-K1) were infected with a recombinant *E. coli* Top10-CFA/I bacterial strain. Using immunocytochemistry, we observed the CFA/I strain attached more to CHO-Le<sup>a</sup> (Fig. 2a) than to CHO-Le<sup>b</sup> cells (Fig. 2b). Furthermore, when we measured the binding of the recombinant CFA/I strain using a quantitative immunofluorescence assay (Fig. 2c), we observed approximately three-fold, significantly higher, bacterial binding to the CHO-Le<sup>a</sup> cell line than to the CHO-Le<sup>b</sup> ( $P = 0.0065$ ), CHO-CP55 and CHO-K1 cell lines ( $P = <0.0001$ ).

**The ETEC CFA/I major subunit CfaB binds more to CHO-Le<sup>a</sup> cells than the ETEC CFA/I minor subunit CfaE.** It has previously been reported that CFA/I has two distinct binding activities with CfaE binding to receptors of unknown structures on the surface of erythrocytes and intestinal epithelial and cultured cells, whilst the major subunit CfaB binds to various glycolipids present on human small intestinal tissue<sup>4,7,16</sup>. To assess if either the major CfaB subunit or the minor CfaE subunit are responsible for the binding of CFA/I to Le<sup>a</sup>, we infected our glyco-engineered cell lines with a CFA/I recombinant strain without the minor subunit CfaE (Top10-CFA/I $\Delta$ E). Using our semi-quantitative immunofluorescence assay, we found that the CFA/I $\Delta$ E strain bound at a significantly higher percentage to the CHO-Le<sup>a</sup> cell line compared to the CHO-Le<sup>b</sup> (Fig. 3a,  $P = 0.0376$ ), CHO-CP55 and CHO-K1 cell lines (Fig. 3a,  $P = <0.0001$ ).

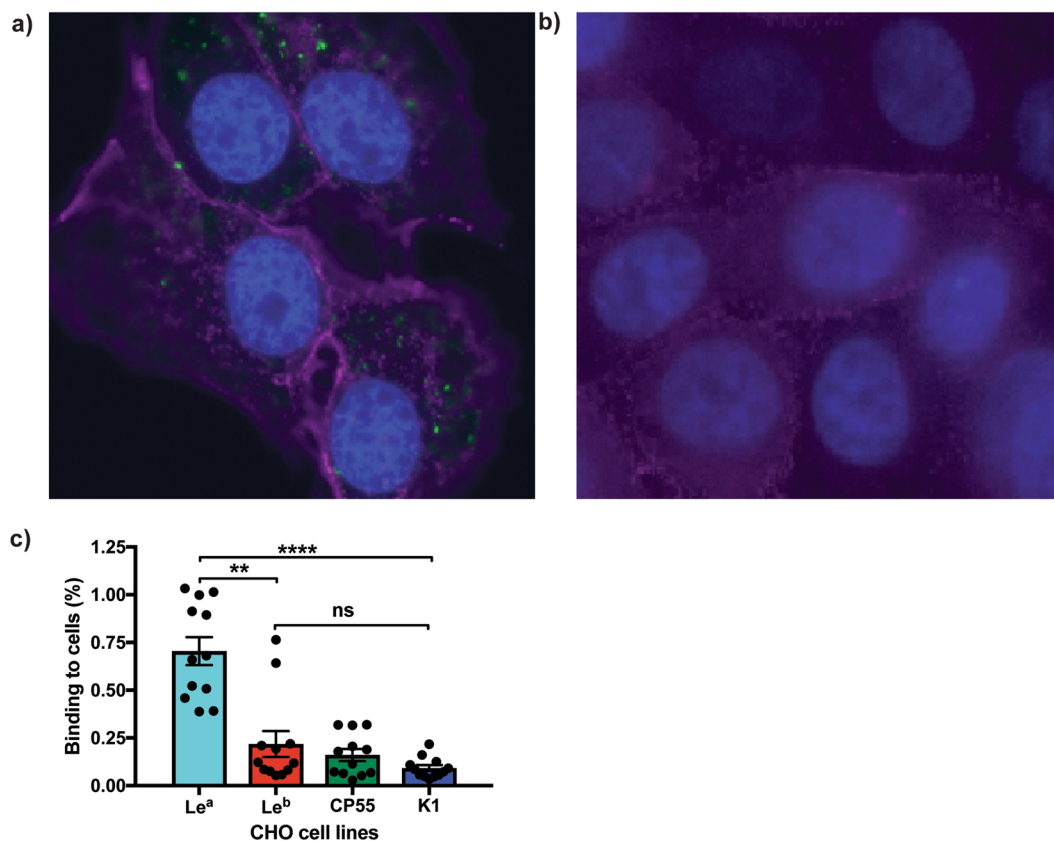
To substantiate these observations, we next performed an inhibition assessment assay using anti-CfaB and anti-CfaE antibodies, the Top10-CFA/I bacterial strain and CHO-Le<sup>a</sup> cells. We found that pre-incubation of the TOP10-CFA/I strain with the anti-CfaB antibody alone, or in combination with equal amounts of the anti-CfaE antibody, significantly reduced the binding of the Top10-CFA/I strain to the CHO-Le<sup>a</sup> cells (Fig. 3b,  $P = <0.0001$ ).



**Figure 1.** Glyco-engineered CHO-Le<sup>a</sup> and CHO-Le<sup>b</sup> cell lines express Lewis antigens. **(a)** SNFG (Symbol Nomenclature for Glycans) diagrams of the Le<sup>a</sup> and Le<sup>b</sup> expressing glycan structures that were engineered to CHO-K1 cell lines for this study. For further details please also see Supplementary Fig. S1 and Table S1 online. **[(b,c)]** Immunocytochemistry staining of the CHO-Le<sup>a</sup> and CHO-Le<sup>b</sup> cell lines. **(b)** CHO-Le<sup>a</sup> cells stained with anti-Le<sup>a</sup> and anti-Le<sup>b</sup> antibodies. **(c)** CHO-Le<sup>b</sup> cells stained with anti-Le<sup>a</sup> and anti-Le<sup>b</sup> antibodies. Lewis antigens are visualized with Alexa Fluor 488-conjugated antibodies (green) and host cell nuclei with DAPI (blue). Magnification x20. **(d)** SDS-PAGE and Western blot analysis of PSGL-1/mIgG2b proteins expressed in the CHO-CP55 (lane 1), CHO-Le<sup>a</sup> (lane 2) and CHO-Le<sup>b</sup> (lane 3) cell lines. In each lane, 1.5 μg of protein was loaded. Membranes were probed with either anti-PSGL1, anti-Le<sup>a</sup> or anti-Le<sup>b</sup> antibodies followed by an anti-mouse IgG secondary antibody.

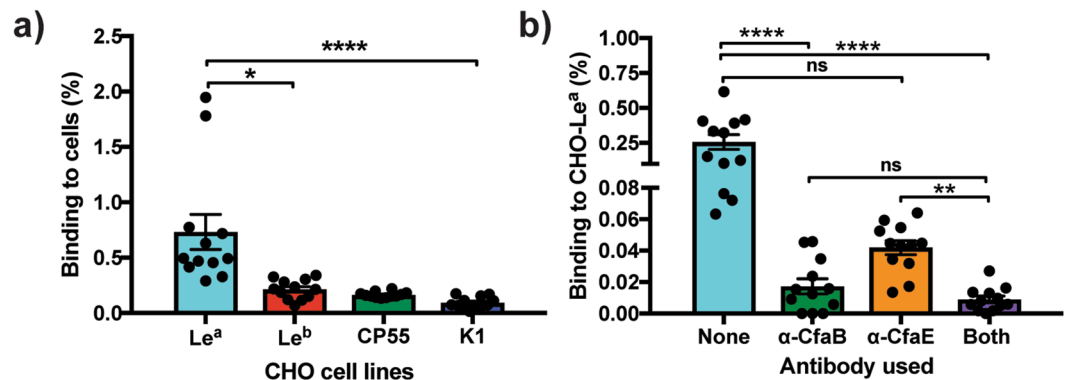
Pose I.D	XP Glide Score	MM-GBSA score	Effective Binding Energy I <sup>a</sup>	Effective Binding Energy II <sup>b</sup>	Ligand solvent GB	Ligand strain energy	Linear Interaction Energy <sup>c</sup>
<b>Selected Le<sup>a</sup>-5 docking poses</b>							
a03	-8.898	-52.583	-0.061	-0.907	-32.956	7.368	-19.627
a04	-10.036	-52.168	-0.061	-0.899	-40.076	15.161	-12.092
a06	-10.182	-49.319	-0.058	-0.850	-37.542	18.196	-11.777
a08	-10.310	-45.768	-0.054	-0.789	-32.258	14.674	-13.509
a09	-8.005	-45.628	-0.053	-0.787	-37.112	10.394	-8.517
<b>Selected Le<sup>b</sup>-6 docking poses</b>							
b01	-8.302	-56.176	-0.056	-0.826	-53.394	0.935	-2.782
b02	-8.798	-55.204	-0.055	-0.812	-36.825	0.081	-18.379
b05	-8.395	-48.488	-0.048	-0.713	-38.728	11.456	-9.759
b07	-8.046	-46.198	-0.046	-0.679	-39.974	5.267	-6.224
b08	-9.352	-44.813	-0.045	-0.659	-41.044	5.922	-3.769

**Table 1.** *In-silico* docking analysis predicts relatively higher binding affinity of the CFA/I docking site to Le<sup>a</sup>-5. Candidate poses selected from the ten highest ranked Le<sup>a</sup>-5 and Le<sup>b</sup>-6 docking scores based on lowest Effective Binding Energy I and II, that predict CfaB binding near/on Lewis antigen moieties of either the Le<sup>a</sup>-5 or Le<sup>b</sup>-6 glycan. All energy score units are measured in kcal/mol. <sup>a</sup>MM-GBSA score/molecular weight of glycan. <sup>b</sup>MM-GBSA score/number of heavy atoms in glycan. <sup>c</sup>MM-GBSA score - Ligand solvent GB.



**Figure 2.** ETEC CFA/I fimbriae bind to the CHO-Le<sup>a</sup> cells. Immunofluorescence staining of (a) CHO-Le<sup>a</sup> and (b) CHO-Le<sup>b</sup> cells infected with TOP10-CFA/I. Infected cell lines were stained with anti-CFA/I antibody (Alexa Fluor 488, green), Le<sup>a</sup> antibody (Alexa Fluor 647, purple), Le<sup>b</sup> antibody (Texas Red, purple) and host nuclei stain (DAPI, Blue). All images are taken at x40 magnification. (c) TOP10-CFA/I bacteria adhere to the CHO-Le<sup>a</sup> cell line more than to CHO-Le<sup>b</sup>, CHO-CP55 and CHO-K1 cells. Graphs represent the percentage of TOP10-CFA/I bacteria binding to each of the cell lines, as measured by quantifiable immunofluorescence analysis. Statistical analysis was performed using ANOVA with Dunn's multiple comparisons test. \*\*Indicates  $P = 0.0065$ , \*\*\*\*Indicates  $P < 0.0001$ . Data presented as Mean and SEM of at least three independent experiments.





**Figure 3.** CfaB, the major subunit of ETEC CFA/I fimbriae, binds to the CHO-Le<sup>a</sup> cells. (a) TOP10-CFA/I ΔE bacteria adhere to the CHO-Le<sup>a</sup> cell line more than to CHO-Le<sup>b</sup>, CHO-CP55 and CHO-K1 cell lines. (b) Inhibition of binding of TOP10-CFA/I bacteria to the CHO-Le<sup>a</sup> cells, using MAbs specific for the CFA/I major subunit CfaB, the minor subunit CfaE, and a mixture of anti-CfaB and anti-CfaE. Graphs represent the percentage of bacteria binding to each of the cell lines, as measured by quantifiable immunofluorescence analysis. Statistical analysis was performed using ANOVA with Dunn's multiple comparisons test. \*Indicates  $P = 0.0376$ , \*\*Indicates  $P = 0.0035$ , \*\*\*\*Indicates  $P = < 0.0001$ . Data presented as Mean and SEM of (a) three independent and (b) two independent experiments.

There was a slight but no significant decrease ( $P = < 0.08$ ) in the binding to CHO-Le<sup>a</sup> cells when the Top10-CFA/I strain was pre-incubated with the anti-CfaE antibody alone, compared to no antibody incubation (Fig. 3b).

**ETEC CFA/I CF family members also bind to CHO-Le<sup>a</sup> cells.** We have previously demonstrated that Bangladeshi children expressing the Le<sup>a</sup> antigen (Le(a + b -), non-secretor phenotype) are more likely to be suffering with symptomatic ETEC infection if they have been infected by ETEC strains expressing CFA/I, CS1, CS2, CS4, CS14 and CS17 CFs<sup>7</sup>. To highlight the flexibility of our glycan models, we also infected them with wild-type ETEC reference strains expressing the above CFs (see materials and methods for further details of the ETEC reference strains used). We observed the wild-type ETEC strain expressing CFA/I bound at a significantly higher percentage to CHO-Le<sup>a</sup> cells, compared to the CHO-Le<sup>b</sup> ( $P = 0.0040$ , Fig. 4), CHO-CP55 and CHO-K1 cells ( $P = < 0.0001$ , Fig. 4). Similarly, CS4 expressing ETEC bacteria attached strongly to Le<sup>a</sup> but not the Le<sup>b</sup> or negative control expressing CHO cells ( $P = < 0.0002$ , Fig. 4). Moreover, the wild-type ETEC strains expressing CS1, CS2, and CS14 adhered significantly more to the CHO-Le<sup>a</sup> than to the CHO-Le<sup>b</sup>, CHO-CP55 and CHO-K1 cells ( $P = < 0.0040$ , Fig. 4). However, we could see no significant difference in the binding of the ETEC strain expressing CS17 to any of our glycan cell line models (Fig. 4).

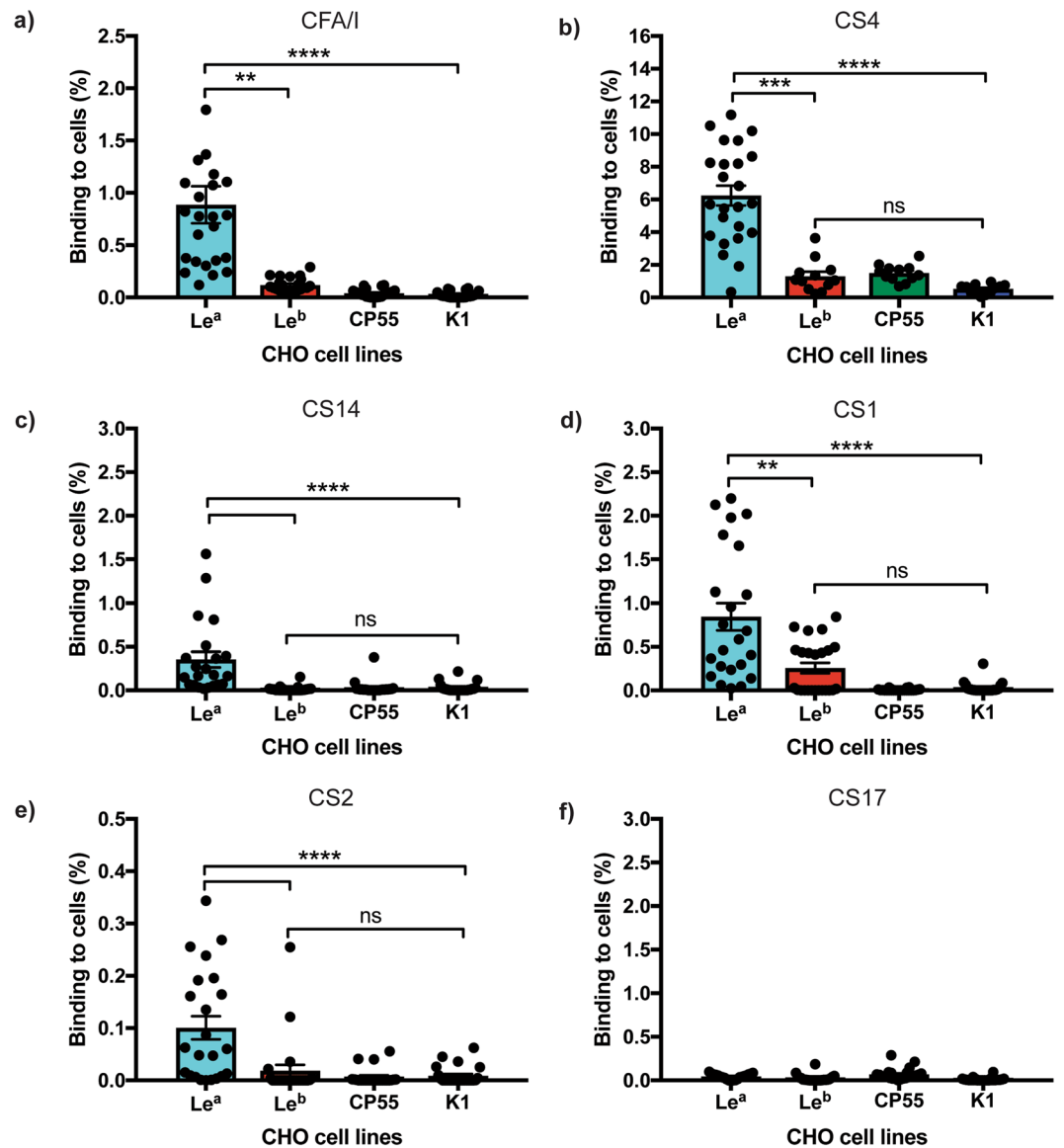
**Molecular docking predicts the Ig-like groove of CfaB is used by ETEC CFA/I to bind to Le<sup>a</sup> expressing glycans.** To provide an insight into why CFA/I fimbriae can bind to Le<sup>a</sup> glycans, we performed molecular docking simulations using pre-existing 3D protein and ligand structures of CFA/I and the Le<sup>a</sup>-5 and Le<sup>b</sup>-6 determinants<sup>9,25-27</sup>. For each glycan, 32 docking simulations were generated around the N terminal Ig-like region of the CfaB major subunit of CFA/I (see Fig. 5c as an example). To predict the most likely of these docking simulations, these poses were ranked based on highest Effective Binding Energy (I and II) (see Supplementary Tables S2 and S3 online, and materials and methods for further details).

Interestingly, we noticed the ten most likely Le<sup>a</sup>-5 poses had relatively higher Effective Binding Energy than the ten most likely Le<sup>b</sup>-6 poses (Fig. 5a and Supplementary Tables S2 and S3 online). Equally, these Le<sup>a</sup>-5 docking poses had on average, higher Linear Interaction Energy than the Le<sup>b</sup>-6 docking poses, suggesting Le<sup>a</sup>-5 interacts more strongly with the CFA/I docking site, compared to Le<sup>b</sup>-6.

Next, we examined if the Ig-like groove region of CfaB could be binding to the Lewis antigen containing regions of our most likely Le<sup>a</sup>-5 and Le<sup>b</sup>-6 docking poses. From this selection, we instantly identified the docking candidate with the highest Effective Binding Energy (I and II) and Linear Interaction Energy to be the Le<sup>a</sup>-5 docking pose, a03 (Table 1). This selection also predicted Thr<sup>29</sup> and up to two other amino acids (Glu<sup>25</sup> Asn<sup>27</sup>) of the CfaB Ig-like groove V<sup>24</sup>EKNITVTASVD<sup>35</sup> region could be binding to the α1,4-FucT (Le<sup>a</sup>) and/or the neighboring β1,3Gal and GalNAc moieties of the Le<sup>a</sup>-5 glycan (Table 1, Fig. 5d, Supplementary Figs S3 and S4 online). In contrast to the most likely Le<sup>b</sup>-6 docking poses (b01, b05, b07 and b08), we observed that only one CfaB amino acid (Thr<sup>29</sup>) might bind to just the α1,2-FucT moiety (Le<sup>b</sup>) of the Le<sup>b</sup>-6 glycan (Table 1, Fig. 5e, Supplementary Figs S3 and S5 online). Furthermore, the comparatively higher ligand strain energy, the lower ligand strain energy scores and relatively high ligand solvent GB scores of Le<sup>a</sup>-5 compared to Le<sup>b</sup>-6 (Table 1), implies less steric hindrance of Le<sup>a</sup>-5 in the CFA/I docking site suggesting that Le<sup>a</sup>-5 is a more distorted structure than the Le<sup>b</sup>-6 glycan.

## Discussion

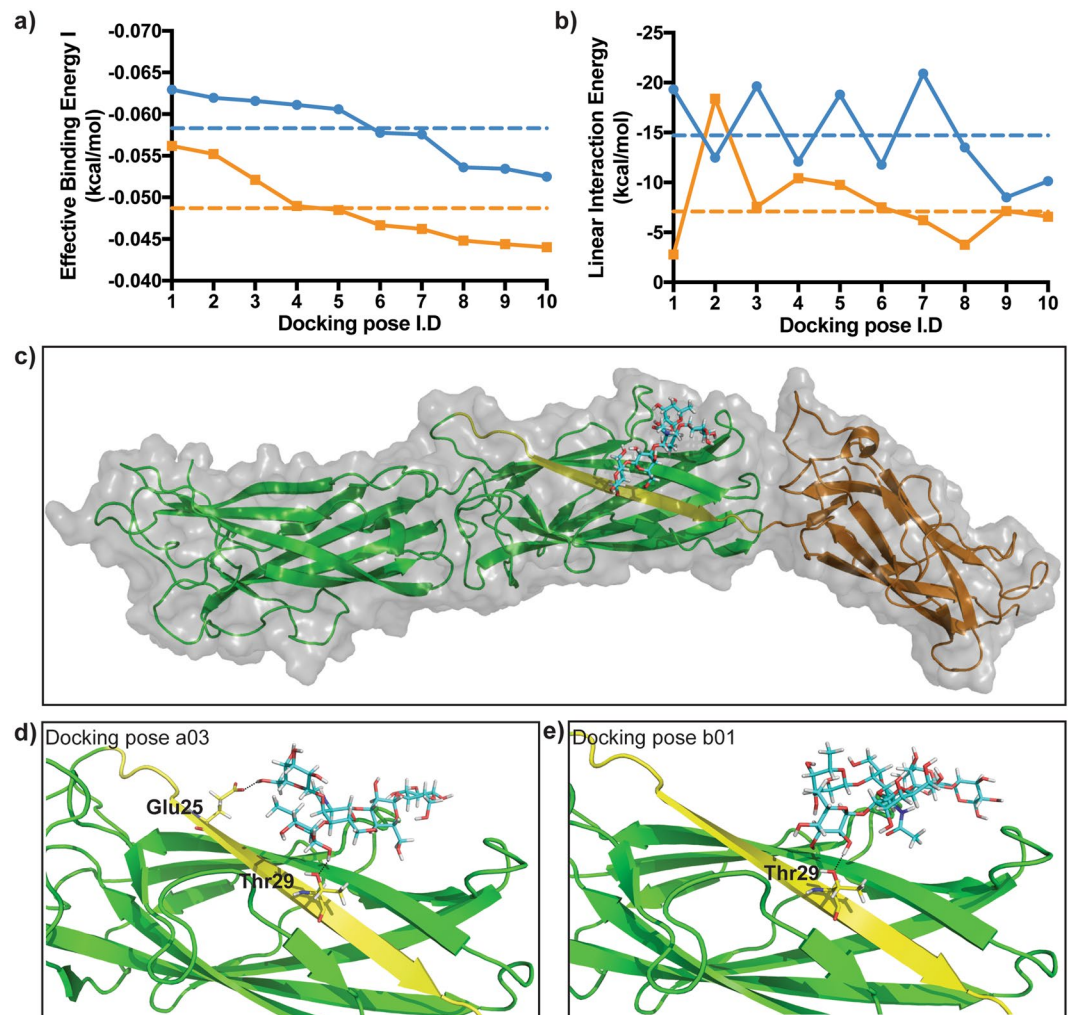
The surface of the mammalian intestinal tract is covered in a rich diversity of mucosal glycans. Such glycans often express genetically defined HBGAs, with HBGA expression varying in the small and large intestine. HBGA glycan expression evolves from birth, and can contribute to intestinal homeostasis, and microbial composition. However, these genetically defined HBGAs can also act as target binding receptors for the virulence factors of



**Figure 4.** ETEC CFA/I and other type 5 pili family members bind to CHO-Le<sup>a</sup> cells. (a) ETEC expressing CFA/I<sup>+</sup> ST<sup>+</sup>/LT<sup>+</sup> (b) ETEC expressing CS4<sup>+</sup>, ST<sup>-</sup>/LT<sup>-</sup> (c) ETEC expressing CS14<sup>+</sup>, ST<sup>+</sup>/LT<sup>-</sup> (d) ETEC expressing CS1<sup>+</sup>, ST<sup>-</sup>/LT<sup>-</sup> (e) ETEC expressing CS2<sup>+</sup>, ST<sup>-</sup>/LT<sup>-</sup>, and (f) ETEC expressing CS17<sup>+</sup>, ST<sup>-</sup>/LT<sup>+</sup> and their binding to CHO-Le<sup>a</sup>, CHO-Le<sup>b</sup>, CHO-CP55 and CHO-K1 cells. Graphs represent the percentage of bacteria binding to each of the cell lines, as measured by quantifiable immunofluorescence analysis. Statistical analysis was performed using ANOVA with Dunn's multiple comparisons test. \*\*Indicates  $P = < 0.0040$ , \*\*\*\*Indicates  $P = 0.0002$ , \*\*\*\*Indicates  $P = < 0.0001$ . Data represent the Mean and SEM of two independent experiments.

microbial pathogens. Such host-pathogen interactions are thought to contribute to pathogen-host species specificity, and pathogen-host tissue tropism<sup>17-19,28</sup>. For example, *FUT2*<sup>-/-</sup> (i.e. HBGA Le<sup>a</sup>-expressing) individuals are genetically immune to certain norovirus genotypes (GI-1, G-II-3 and G-II-4) compared to *FUT2*<sup>+/+</sup> or *FUT2*<sup>+/-</sup> individuals<sup>17</sup>. Similarly *FUT2*<sup>+/+</sup> pigs are more susceptible to porcine pathogenic ETEC strains expressing F18<sup>+</sup> fimbriae<sup>29</sup>.

We have previously published clinical evidence to suggest that ETEC CFA/I uses the HBGA Le<sup>a</sup> as a small intestinal host receptor, as Bangladeshi children expressing this Le<sup>a</sup> antigen (caused by a *FUT2* mutation, non-secretor status) are more susceptible to disease caused by ETEC expressing CFA/I and related CF family members, than are Bangladeshi children expressing the Le<sup>b</sup> antigen (functional *FUT2*, secretor status)<sup>5-7</sup>. Expression of HBGAs (Lewis and ABO(H)) on the surface of the human intestinal mucosa is driven by the *FUT2* and *FUT3* genes, with *FUT2* encoding an  $\alpha$ 1,2-fucosyltransferase ( $\alpha$ 1,2-FucT) and *FUT3* encoding an  $\alpha$ 1,3/4-fucosyltransferase ( $\alpha$ 1,3/4-FucT) (Supplementary Fig. S1 online)<sup>17</sup>. Expression of only *FUT3* results in Le<sup>a</sup> antigen expression, whilst expression of both *FUT3* and *FUT2* results in the expression of Le<sup>b</sup> determinants on small intestinal glycans<sup>17-19</sup>.



**Figure 5.** Molecular docking predicts CfaB can bind to Le<sup>a</sup>-5 and Le<sup>b</sup>-6 glycans. **(a)** Effective Binding Energy I (kcal/mol) of the ten highest ranked Le<sup>a</sup>-5 and Le<sup>b</sup>-6 poses. **(b)** Linear Interaction Energy (kcal/mol) of the ten highest ranked Le<sup>a</sup>-5 and Le<sup>b</sup>-6 poses. Blue lines are Le<sup>a</sup>-5 poses and orange lines are Le<sup>b</sup>-6 poses. Dotted lines are respective mean values for the ten highest ranked Le<sup>a</sup>-5 (blue) and Le<sup>b</sup>-6 (orange) poses analysed. **(c)** Overall surface view and cartoon representation of Le<sup>a</sup>-5 binding to the N-terminal Ig-like groove of the major CFA/I subunit CfaB. **(d)** Cartoon representation of docking pose a03, showing hydrogen bond interaction of residues Glu<sup>25</sup> and Thr<sup>29</sup> of the CfaB Ig-like groove with Le<sup>a</sup>-5. **(e)** Cartoon representation of docked complex b01, showing hydrogen bond interaction of Thr<sup>29</sup> of the CfaB Ig-like groove with Le<sup>b</sup>-6. For **(c)** to **(e)** the Le<sup>a</sup>-5 and Le<sup>b</sup>-6 ligands are stick representations (cyan) with atoms interacting with CfaB coloured in yellow and hydrogen bonds displayed as black dotted lines. For the CFA/I fimbriae, the minor subunit CfaE is the green ribbon, the major subunit CfaB is the dark yellow ribbon and the CfaB N terminal Ig-like groove is the bright yellow ribbon.

To create a defined glycan model of the human small intestine to study the binding of ETEC CFA/I fimbriae to Lewis antigens, we glyco-engineered the well-defined and naturally HBGA devoid CHO-K1 cell line<sup>22</sup>. To generate CHO-K1 cells carrying Le<sup>a</sup> or Le<sup>b</sup> determinants on their cell surface, we expressed the extended core 1 glycan (GlcNAc $\beta$ 3Gal $\beta$ 3GalNAc $\alpha$ ) enzyme B3GNT3, the type 1 chain glycan (Gal $\beta$ 3GlcNAc)-encoding enzyme B3GALT5, and the Lewis gene-encoding enzyme FUT3 alone (generating Le<sup>a</sup>) or together with the H gene-encoding FUT1 (generating Le<sup>b</sup>) on a PSGL-1/mIgG2b fusion protein carrying probe (See Supplementary Fig. S1 and Table S1 online for further details)<sup>20–24</sup>. Expression of the B3GALT5 enzyme facilitates the biosynthesis of type 1 chains (Gal $\beta$ 3GlcNAc) on both N- and O-glycans (on the latter following co-expression of the extended core 1 enzyme B3GNT3), which is the obligate precursor for the Lewis (FUT3) and H-gene (FUT1) enzymes. This enabled us to engineer recombinant CHO-K1 cell lines carrying abundant Le<sup>a</sup> or Le<sup>b</sup> antigen substitutions<sup>24</sup>.

Initially, we evaluated the binding specificity of ETEC CFA/I fimbriae to our stable Lewis antigen-expressing cell lines, by infecting them with recombinant bacteria expressing CFA/I fimbriae. We found the CFA/I expressing strain attached more to the CHO-Le<sup>a</sup> cell line, than to the CHO-Le<sup>b</sup> or the Lewis antigen negative control cell lines (CHO-CP55 and CHO-K1). This may suggest that the  $\alpha$ 1,2-linked fucose of  $\alpha$ 1,2-FucT which is used to create Le<sup>b</sup> could be either blocking the CFA/I binding sites on the Le<sup>a</sup> receptor, or making the binding sites less accessible for the CFA/I fimbriae and thus preventing ETEC CFA/I attachment.

To define which subunit of ETEC CFA/I fimbriae are responsible for the binding of CFA/I fimbriae to our CHO-Le<sup>a</sup> cell line, we also infected our cell lines with a recombinant *E.coli* strain expressing CFA/I without the minor subunit (Top10-CFA/I $\Delta E$ ). As expected<sup>7</sup>, the *CfaE* deleted strain adhered more to CHO-Le<sup>a</sup> cells than to CHO-Le<sup>b</sup>, CHO-CP55 or CHO-K1 cells. Similarly, pre-incubation of the Top10-CFA/I recombinant strain with either anti-CfaB or an equal mix of anti-CfaB and CfaE antibodies, significantly reduced CFA/I binding to CHO-Le<sup>a</sup> cells. However, the bacterial binding after the pre-incubation of Top10-CFA/I with only anti-CfaE compared to no antibody pre-incubation was reduced, but not statically significant. As mature CFA/I fimbriae contain >1,000 copies of CfaB and one or a few copies of CfaE<sup>4</sup>, we therefore conclude the major subunit CfaB is naturally the more dominant subunit for CFA/I binding to small intestinal Le<sup>a</sup> glycans.

To help understand why the CfaB subunit of ETEC CFA/I fimbriae binds to Le<sup>a</sup> but not Le<sup>b</sup> glycans, we performed computational molecular docking analysis. Our selected CFA/I docking site encompassed the amino acids V<sup>24</sup>EKNITVTASVD<sup>35</sup> of the highly conserved CfaB Ig-like groove, found in major subunits of ETEC CFA/I related CFs (Supplementary Figs S2 and S5 online) and class 1 pili of bacteria that can cause urinary and respiratory infections by binding to host glycolipids containing HBGAs<sup>9,17</sup>.

Supporting our binding studies using glycan-defined CHO-K1 cells, as well as previous clinical observations<sup>6,7</sup>, our *in-silico* defined CFA/I docking site bound with a relatively higher binding affinity and higher interaction preference to the Le<sup>a</sup> than to Le<sup>b</sup> glycan. Moreover, the ligand strain energy as well as the small size of Le<sup>a</sup>-5, might explain its relative distortion in the CFA/I binding site as compared to that of the Le<sup>b</sup>-6 glycan. We hypothesise this relative distortion, could be one reason why multiple amino acids (Asn<sup>27</sup>, Thr<sup>29</sup> and Glu<sup>25</sup>) of the CfaB Ig-like fold region (Supplementary Figs S2 and S5 online), docked to the  $\alpha$ 1,4-linked fucose and/or several surrounding moieties of the Le<sup>a</sup>-5 glycan. We now plan to perform more computational simulations as well as further *in-vitro* analysis to assess these observations further.

To highlight the flexibility of our small intestinal glycan defined like model, we also infected our Lewis expressing CHO-K1 cell lines with wild-type ETEC reference strains expressing CFA/I as well as CFA/I related CFs; the latter having been suggested to also bind to the small intestinal mucosa of children expressing Le<sup>a</sup><sup>7</sup>. As expected, the wild-type ETEC strain expressing CFA/I bound significantly more to CHO-Le<sup>a</sup> cells than to the Le<sup>b</sup> or negative cell lines. Similarly, the ETEC strains expressing the related CFs CS1, CS2, CS4, CS14, but not CS17, bound more to CHO-Le<sup>a</sup> cells.

Interestingly, the amino acids that we hypothesise by molecular docking to be important for CfaB binding to Le<sup>a</sup>-5, are also highly conserved in the major subunits of the related CFA/I CFs we have studied (see Supplementary Fig. S6 online for more details). However, the binding capacity differences of these wild-type ETEC strains expressing these CFs to CHO-Le<sup>a</sup> cells may be due to evolved conformational differences amongst family members (Supplementary Fig. S6 online)<sup>4</sup>. Therefore, one reason why the CS17 ETEC strain does not bind to CHO-Le<sup>a</sup> cells is that amino acid Arg<sup>31</sup>, located only a few positions away from the polar Asn<sup>27</sup> and Thr<sup>29</sup> amino acids (R group, Supplementary Fig. S6 online), is potentially leading to CS17 CF structural and conformational changes near these predicted binding sites.

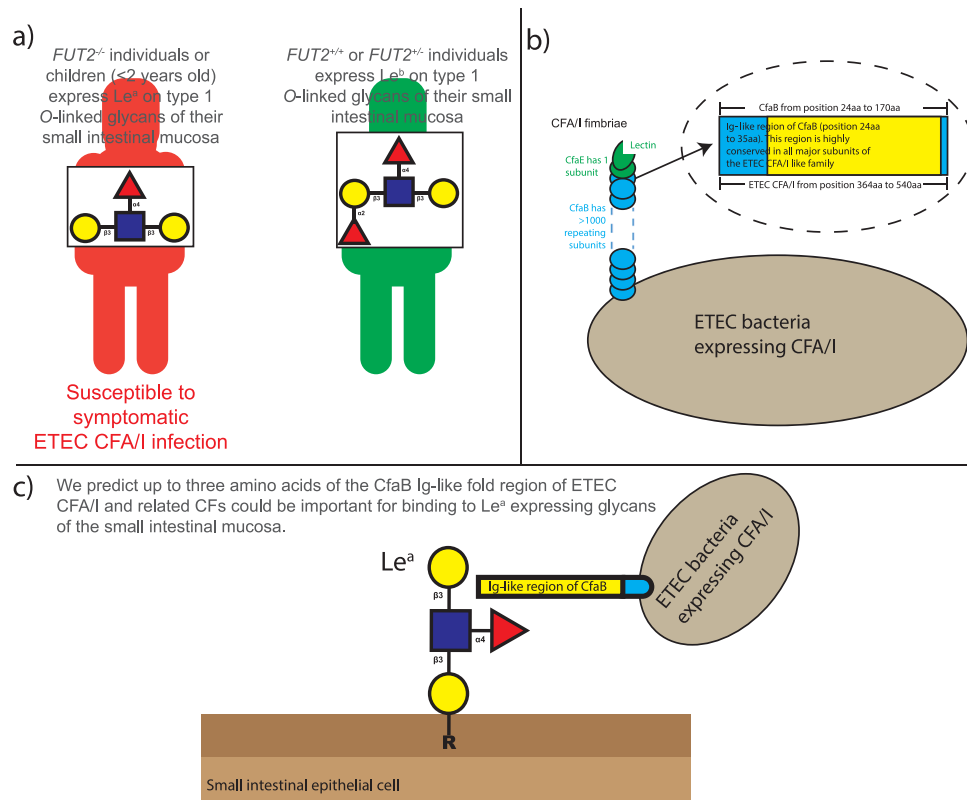
A critical first step in a microbial pathogenesis is frequently the attachment to host cell glycans. In particular, targets on host tissues include the ubiquitously expressed HBGAs of different mucosae<sup>17–19,28</sup>. Structural and functional studies are now starting to reveal insights into why individuals expressing different HBGAs are at an increased risk of infections such as those caused by *Vibrio cholera*, *Pseudomonas aeruginosa*, norovirus, rotavirus and *Helicobacter pylori*<sup>17,28</sup>. By developing Lewis antigen cell-based models of the human small intestine, as well as performing docking studies, we have further defined why HBGA Le<sup>a</sup> – expressing non-secretor (*FUT2*<sup>-/-</sup>) individuals, or young children (<2 years of age) are more susceptible to ETEC expressing CFA/I, or other related CFs family members<sup>5,6,8,28</sup>. Subsequently, our understanding and characterisation of these host-pathogen binding patterns could represent a critical point at which the adherence of ETEC expressing CFA/I and related CF family members, can be prevented with vaccines and/or anti-infection therapeutics that block this interaction (Fig. 6).

## Materials and Methods

**Plasmids used to construct glyco-engineered cell lines.** Plasmids encoding the P-selectin glycoprotein-1/mouse immunoglobulin IgG2b Fc fragment (PSGL-1/mIgG2b) fusion protein, the human extended core 1 (GlcNAc $\beta$ 3Gal $\beta$ 3GalNAc $\alpha$ ) enzyme (B3GNT3), the type 1 chain (Gal $\beta$ 3GlcNAc) enzyme (B3GALT5), the Lewis gene  $\alpha$ 3/4-fucosyltransferase (FUT3) and the blood group H gene-encoded  $\alpha$ 2-fucosyltransferase (FUT1) were constructed as described previously (See Supplementary Table S1 online for details of the constructed plasmids)<sup>20–22,24</sup>. The pCMV/FUT1/Zeo rather than a FUT2 expression plasmid was selected to be used in this study, as it has been shown to better support Le<sup>b</sup> expression in CHO cells<sup>24</sup>.

**Construction of glyco-engineered cell lines.** Adherent CHO-K1 cells (ATCC<sup>®</sup>, Manassas, VA, USA) were seeded in six-well cell culture plates containing Dulbecco's Modified Eagle's medium (DMEM, Lonza), supplemented with 10% fetal bovine serum (FBS, Invitrogen, Waltham, MA, USA). All transfection experiments were performed 24 hours after seeding (70–80% cellular confluency). Cellular transfection was performed in accordance with manufacturer's instructions using Lipofectamine 2000 (Invitrogen). CHO-K1 cells were transfected with the pEF1 $\alpha$ /PSGL-1/mIgG2b/PAC plasmid to create the CHO-CP55 cell line. CHO-CP55 cells were then co-transfected with a cocktail of plasmids; pCMV/C1- $\beta$ 1,3GlcNAcT/Neo (human extended core 1 enzyme), pCMV/GalT5/Gpt (type 1 chain enzyme), in combination with of pCMV/FUT3/Hyg for Le<sup>a</sup> expression (CHO-Le<sup>a</sup> cell line), and combined with pCMV/FUT1/Zeo for Le<sup>b</sup> expression (CHO-Le<sup>b</sup> cell line) (Supplementary Fig. S1 and Table S1 online). To generate stable transfectants, all the five expression vectors were linearized with *Avr* II (New England BioLabs, Ipswich, MA, USA), and equal concentrations of plasmids encoding glycosyltransferases were used.





**Figure 6.** Working hypothesis. **(a)** *FUT2*<sup>-/-</sup> individuals or children (<2 years old) express Le<sup>a</sup> on type 1 O-linked glycans of the small intestinal mucosa. We have evidence to suggest they are susceptible to symptomatic ETEC CFA/I (and related CFA/I CF family members) infection. *FUT2*<sup>+/+</sup> or *FUT2*<sup>+/-</sup> individuals (>2 years old) express Le<sup>b</sup> on type 1 O-linked glycans of the small intestinal mucosa. We have not found these individuals to not be susceptible to ETEC CFA/I (and related CFA/I CF family members) infection. **(b)** ETEC CFA/I fimbriae contains >1,000 copies of a major subunit (CfaB), with one or a few copies of the tip residing minor subunit (CfaE). A 12-aa stretch V<sup>24</sup>EKNITVTASVD<sup>35</sup> of the Ig-like binding groove region of CfaB served as a CFA/I docking site our *in-silico* docking analysis. This 12 amino acid stretch also shares structural similarities with all ETEC CF major subunits of the CFA/I CF like family, and with class 1 pili from bacteria that can cause urinary and respiratory infections by binding to host glycolipids containing HBGAs. **(c)** Using our small intestinal like glycan CHO-K1 cell lines, we have demonstrated that CfaB of ETEC CFA/I fimbriae, as well as four related CFs, bind more to our CHO-K1 cell-line expressing Le<sup>a</sup>, compared to cells carrying Le<sup>b</sup> or the CHO-K1 wild-type glycan phenotype. Using *in-silico* docking analysis, we predict up to three amino acids (Glu<sup>25</sup>, Asn<sup>27</sup>, Thr<sup>29</sup>) found in the immunoglobulin (Ig)-like groove region of CfaB of CFA/I and related CF fimbriae, could be important for the preferential and higher affinity binding of CFA/I fimbriae to Le<sup>a</sup> glycans. These findings may lead to a better molecular understanding of ETEC pathogenesis, aiding in the development of vaccines and/or anti-infection therapeutics, which block such host-pathogen interactions.

Following transfection (48 hours), all the transfected cell lines were incubated in selection medium as stated in the Supplementary Materials and Methods online. Two weeks after transfection, drug resistant clones were picked and transferred to 96-well plates containing their corresponding selection medium and propagated. Le<sup>a</sup> and Le<sup>b</sup> expression was assessed by immunocytochemistry and Western blot using monoclonal antibodies specific for the Le<sup>a</sup> and Le<sup>b</sup> determinants (see Supplementary Materials and Methods online for further details). Two clones-CHO-PSGL1-le<sup>a</sup>-a1 and CHO-PSGL1-le<sup>b</sup>-b4 (Fig. 1) were selected, expanded and used for ETEC cell binding experiments.

**Bacterial strains used in infection experiments.** The previously constructed recombinant Top10-CFA/I (Amp<sup>R</sup>) and Top10-CFA/I/E<sup>-</sup> (Cm<sup>R</sup>) strains were cultured as previously described<sup>7</sup>. The following wild-type ETEC strains and natural mutants were also used: E3006 (258909-3; CFA/I<sup>+</sup> ST<sup>+</sup>/LT<sup>+</sup>), E120 (60R936; CS1<sup>+</sup>, ST<sup>-</sup>/LT<sup>-</sup>), E3017 (58R957; CS2<sup>+</sup>, ST<sup>-</sup>/LT<sup>-</sup>), E3037 (62R486; CS4<sup>+</sup>, ST<sup>-</sup>/LT<sup>-</sup>), E3013 (CS14<sup>+</sup>, STh<sup>+</sup>/LT<sup>-</sup>), and E3014 (CS17<sup>+</sup>, ST<sup>-</sup>/LT<sup>+</sup>)<sup>4,30</sup>. All ETEC strains were cultured as previously described<sup>7,30</sup>. Expression of CFs was tested by agglutination assays using monoclonal CF antibodies<sup>4,30</sup>.

In preparation for infection experiments, bacterial cultures were centrifuged at 10,000 × g for 10 minutes. The bacterial pellets were then washed and re-suspended in PBS before being re-centrifuged. Recombinant strains were re-suspended in PBS and ETEC strains were re-suspended in 0.5% D-mannose/PBS solution at an OD 0.8/ mL (680 nm) to block possible binding by Type I fimbriae.

**Infection of CHO cell lines.** Each cell line was seeded at a concentration of  $0.5 \times 10^5$ /mL into microscope well slides (open  $\mu$ -Slide, Ibidi). Three days after seeding, the well culture media was removed, washed with warm PBS and then replaced with 275  $\mu$ L of DMEM with 1% FBS and 2nM L-glutamine in preparation for infection experiments. Cell lines were infected in duplicate with 25  $\mu$ L of the corresponding bacterial suspensions ( $\sim 25$  MOI) for three hours in a humidified incubator at 37°C and 5% CO<sub>2</sub>. Following three washes with PBS, cells were fixed in paraformaldehyde for 10 minutes before being washed a further two times with PBS. Fixed cells were stained by immunofluorescence (see below) to determine the number of adherent bacteria to the cell lines.

**Inhibition of binding of the Top10-CFA/I ETEC strain to engineered CHO cells using monoclonal antibodies.** The monoclonal anti-CfaB and CfaE antibodies were a kind gift from Professor Weiping Zhang of Kansas State University, USA. The CHO-Le<sup>a</sup> cell line and the recombinant Top10-CFA/I strain were prepared as described above. A suspension (0.8OD/mL) of the Top10-CFA/I strain was mixed with either a 1:50 dilution of (i) anti-CfaB, (ii) anti-CfaE, (iii) a mixture of anti CfaB and anti-CfaE or (iv) no antibody mix. The bacterial/antibody mixtures were then mixed thoroughly and incubated for 30 minutes at room temperature. Cell lines were then infected with the bacterial mixtures as described above.

**Immunofluorescence staining of infected cell lines and quantifiable immunofluorescence analysis.** All microscopy was performed using the inverted LSM700 confocal microscope with Zeiss Zen Blue software. Infected cells were stained for immunofluorescence microscopy using the protocol stated in the Supplementary Materials and Methods online.

To quantify bacterial adhesion, bacteria were visualised using the Alexa Fluor 488 channel and the CHO cells with the DAPI channel. All immunofluorescence settings remained the same for each experiment performed. Using the Multiple Single Positions (Tiles) position array tool available in the Zen Blue microscope software, six random tiles of each microscopic slide chamber were taken at x20 magnification. For each experiment performed, a total of two chambers per CHO cell line and infection were imaged (i.e. 12 random tile scans per cell line and bacterial infection). Confocal images in Zen format were analysed using the Velocity 3D image software (PerkinElmer, California, USA). Bacterial adhesion was quantified by measuring the mean surface area of bacteria (Alexa Fluor 488 channel fluorescence,  $\mu\text{m}^2$ ) and the mean surface area of CHO cells (DAPI channel fluorescence,  $\mu\text{m}^2$ ). Binding of bacteria to CHO cell lines was expressed as a percentage of the total mean surface area of bacteria divided by the total mean surface area of CHO cells multiplied by 100.

**Molecular docking.** *In-silico* docking studies were performed using the Glide module<sup>31</sup> version 6.2, in Schrödinger, LLC, New York, NY, 2017. The X-ray crystal structure of the fused complex containing the CfaE and CfaB subunits of ETEC CFA/I fimbriae at 2.3 Å was obtained from the RSCB Protein Data Bank (PDB ID: 3F83)<sup>9,25,32</sup>, and prepared for molecular docking using the Protein Preparation Wizard<sup>33</sup> of the Maestro, version 10.5 of the Schrödinger, LLC, New York, NY, 2017 software. Le<sup>a</sup>-5 (CID:11051152) and Le<sup>b</sup>-6 (CID:91852492) structures were downloaded from PubChem<sup>26,27</sup>, and converted to 3D before being processed for *in-silico* experiments using the Ligprep module<sup>34</sup> of the Schrödinger, Maestro v10.5 software. This module generated a number of conformers (32 conformers) for each structure based on various tautomers, stereochemistries, ionization states and checking various ring conformations at a pH range set between  $7 \pm 2$ , followed by energy minimization with OPLS-2005 force field<sup>31</sup>.

Molecular Docking was performed using the Receptor Grid Generation panel module of the Glide, Schrödinger, LLC, New York, NY, 2017 software. A 12-aa stretch V<sup>24</sup>EKNITVTASVD<sup>35</sup> of the Ig-like binding groove region of CfaB was used to a build grid and served as our CFA/I docking site<sup>9</sup>. The extra precision (XP) method of Glide dock was used for the docking experiments of Le<sup>a</sup>-5 and Le<sup>b</sup>-6 conformers to the CFA/I docking site with the sampling of the ligand being kept flexible during docking<sup>35</sup>.

The relative binding affinity score of our CFA/I docking complex to Le<sup>a</sup>-5 and Le<sup>b</sup>-6 was calculated (based on the XP docked complexes) using the MM-GBSA method (Molecular mechanics with generalized born surface area), available in the Schrödinger's tool Prime software (Schrödinger, LLC, New York, NY, 2017)<sup>36,37</sup>. Apart from the MM-GBSA ( $\Delta G$ ) energy, a few other parameters were studied to further elucidate the docking results. These include: Effective Binding Energy (see calculations (1) and (2) below) which rescales the binding affinity of the Ligand strain energy to represent the extent of ligand distortion, ligand solvation energy (Lig solv GB) to represent the binding energy of ligand in solvent, and Linear Interaction Energy to represent the extent of binding affinity of ligand towards protein over solvent (see calculation (3) below).

- (1) Effective Binding Energy I = MM-GBSA/molecular weight of ligand  
Molecular weight of Le<sup>a</sup>-5 ligand = 853.774  
Molecular weight of Le<sup>b</sup>-6 ligand = 999.916
- (2) Effective Binding Energy II = MM-GBSA/number of heavy atoms in ligand  
Number of heavy atoms in Le<sup>a</sup>-5 ligand = 58  
Number of heavy atoms in Le<sup>b</sup>-6 ligand = 68
- (3) Linear Interaction Energy = MM-GBSA – Ligand solvent GB.

Interesting docking sites were selected based on Effective Binding Energy (I and II), the number of interactions from the conserved region of CfaB, and docked complexes interacting with the Lewis antigen or neighbouring moieties of either Le<sup>a</sup>-5 or Le<sup>b</sup>-6.

**Statistical analysis.** Binding of bacteria to the different CHO cell lines were expressed as mean percentages of added bacteria binding to the cells in 12 confocal tile scans. Inhibition results were obtained from two

independent experiments performed in duplicate. All other binding experiments were obtained from at least three independent experiments performed in duplicate. Statistical analysis was performed using Graph Pad Prism 6. ANOVA using Dunn's multi comparisons test was used to compare the mean percentage of bacteria binding to the CHO-Le<sup>a</sup>, CHO-Le<sup>b</sup>, CHO-CP55 and CHO-K1 cell lines. For additional verification, the Wilcoxon matched pairs sign rank test was used to compare the mean percentage of bacteria binding to the CHO-Le<sup>a</sup> cell line compared to CHO-Le<sup>b</sup> cell line. Significance was set at a P value of <0.05.

**Data availability.** The raw datasets generated during and/or analysed during the current study are available in the Figshare repository, (<https://figshare.com/s/41a7b658f9474b6dfd53>).

## References

- Bourgeois, A. L., Wierzbza, T. F. & Walker, R. I. Status of vaccine research and development for enterotoxigenic Escherichia coli. *Vaccine* **34**, 2880–2886, <https://doi.org/10.1016/j.vaccine.2016.02.076> (2016).
- Isidean, S. D., Riddle, M. S., Savarino, S. J. & Porter, C. K. A systematic review of ETEC epidemiology focusing on colonization factor and toxin expression. *Vaccine* **29**, 6167–6178, <https://doi.org/10.1016/j.vaccine.2011.06.084> (2011).
- Torres, A. G., Zhou, X. & Kaper, J. B. Adherence of diarrheagenic Escherichia coli strains to epithelial cells. *Infect Immun* **73**, 18–29, <https://doi.org/10.1128/IAI.73.1.18-29.2005> (2005).
- Madhavan, T. P. & Sakellaris, H. Colonization factors of enterotoxigenic Escherichia coli. *Adv Appl Microbiol* **90**, 155–197, <https://doi.org/10.1016/bs.aambs.2014.09.003> (2015).
- Mottram, L., Wiklund, G., Larson, G., Qadri, F. & Svennerholm, A. M. FUT2 non-secretor status is associated with altered susceptibility to symptomatic enterotoxigenic Escherichia coli infection in Bangladesh. *Sci Rep* **7**, 10649, <https://doi.org/10.1038/s41598-017-10854-5> (2017).
- Ahmed, T. *et al.* Children with the Le(a+ b–) blood group have increased susceptibility to diarrhea caused by enterotoxigenic Escherichia coli expressing colonization factor I group fimbriae. *Infect Immun* **77**, 2059–2064, <https://doi.org/10.1128/IAI.01571-08> (2009).
- Jansson, L., Tobias, J., Lebens, M., Svennerholm, A. M. & Teneberg, S. The major subunit, CfaB, of colonization factor antigen i from enterotoxigenic Escherichia coli is a glycosphingolipid binding protein. *Infect Immun* **74**, 3488–3497, <https://doi.org/10.1128/IAI.02006-05> (2006).
- Kononova, S. V. How Fucose of Blood Group Glycotypes Programs Human Gut Microbiota. *Biochemistry (Mosc)* **82**, 973–989, <https://doi.org/10.1134/S0006297917090012> (2017).
- Li, Y. F. *et al.* Structure of CFA/I fimbriae from enterotoxigenic Escherichia coli. *Proc Natl Acad Sci USA* **106**, 10793–10798, <https://doi.org/10.1073/pnas.0812843106> (2009).
- Madhavan, T. P., Riches, J. D., Scanlon, M. J., Ulett, G. C. & Sakellaris, H. A. Binding of CFA/I Pili of Enterotoxigenic Escherichia coli to Asialo-GM1 is Mediated by the Minor Pilin, CfaE. *Infect Immun* **84**(5), 1642–1649, <https://doi.org/10.1128/IAI.01562-15> (2016).
- Rudin, A., McConnell, M. M. & Svennerholm, A. M. Monoclonal antibodies against enterotoxigenic Escherichia coli colonization factor antigen I (CFA/I) that cross-react immunologically with heterologous CFAs. *Infect Immun* **62**, 4339–4346 (1994).
- Lásaro, M. O. *et al.* Combined Vaccine Regimen Based on Parenteral Priming with a DNA Vaccine and Administration of an Oral Booster Consisting of a Recombinant Salmonella enterica Serovar Typhimurium Vaccine Strain for Immunization against Infection with Human-Derived Enterotoxigenic Escherichia coli Strains. *Infect Immun* **72**(11), 6480–6491, <https://doi.org/10.1128/IAI.72.11.6480-6491.2004> (2004).
- Ranallo, R. T., Fonseka, C. P., Cassels, F., Srinivasan, J. & Venkatesan, M. M. Construction and Characterization of Bivalent Shigella flexneri 2a Vaccine Strains SC608(pCFAI) and SC608(pCFAI/LTB) That Express Antigens from Enterotoxigenic Escherichia coli. *Infect Immun* **73**(1), 258–267, <https://doi.org/10.1128/IAI.73.1.258-267.2005> (2005).
- Rudin, A., Olbe, L. & Svennerholm, A. M. Monoclonal antibodies against fimbrial subunits of colonisation factor antigen I (CFA/I) inhibit binding to human enterocytes and protect against enterotoxigenic Escherichia coli expressing heterologous colonization factors. *Microbial Pathogenesis* **20**, 35–45 (1996).
- Rudin, A. & Svennerholm, A. M. Identification of a cross-reactive continuous B-cell epitope in enterotoxigenic Escherichia coli colonisation factor antigen I. *Infect Immun* **64**(11), 4508–4513 (1996).
- Baker, K. K., Levine, M. M., Morison, J., Phillips, A. & Barry, E. M. CfaE tip mutations in enterotoxigenic Escherichia coli CFA/I fimbriae define critical human intestinal binding sites. *Cell Microbiol* **11**, 742–754, <https://doi.org/10.1111/j.1462-5822.2009.01287.x> (2009).
- Cooling, L. Blood Groups in Infection and Host Susceptibility. *Clin Microbiol Rev* **28**, 801–870, <https://doi.org/10.1128/CMR.00109-14> (2015).
- Bjork, S., Breimer, M. E., Hansson, G. C., Karlsson, K. A. & Leffler, H. Structures of blood group glycosphingolipids of human small intestine. A relation between the expression of fucolipids of epithelial cells and the ABO, Le and Se phenotype of the donor. *J Biol Chem* **262**, 6758–6765 (1987).
- Finne, J. *et al.* Novel polyfucoylated N-linked glycopeptides with blood group A, H, X, and Y determinants from human small intestinal epithelial cells. *J Biol Chem* **264**, 5720–5735 (1989).
- Liu, J., Gustafsson, A., Breimer, M. E., Kussak, A. & Holgersson, J. Anti-pig antibody adsorption efficacy of {alpha}-Gal carrying recombinant P-selectin glycoprotein ligand-1/immunoglobulin chimeras increases with core 2 {beta}1, 6-N-acetylglucosaminyltransferase expression. *Glycobiology* **15**, 571–583, <https://doi.org/10.1093/glycob/cwi037> (2005).
- Liu, J. & Holgersson, J. Recombinant Galalpha1,3Gal-substituted mucin/immunoglobulin chimeras: a superior absorber of anti-pig antibodies. *Transplant Proc* **32**, 859 (2000).
- Liu, J., Jin, C., Cherian, R. M., Karlsson, N. G. & Holgersson, J. O-glycan repertoires on a mucin-type reporter protein expressed in CHO cell pools transiently transfected with O-glycan core enzyme cDNAs. *J Biotechnol* **199**, 77–89, <https://doi.org/10.1016/j.jbiotec.2015.02.017> (2015).
- Liu, J., Qian, Y. & Holgersson, J. Removal of xenoreactive human anti-pig antibodies by absorption on recombinant mucin-containing glycoproteins carrying the Galalpha1,3Gal epitope. *Transplantation* **63**, 1673–1682 (1997).
- Holgersson, J. & Lofling, J. Glycosyltransferases involved in type 1 chain and Lewis antigen biosynthesis exhibit glycan and core chain specificity. *Glycobiology* **16**, 584–593, <https://doi.org/10.1093/glycob/cwj090> (2006).
- The Protein Databank. Structure of fusion complex of the minor pilin CfaE and major pilin CfaB of CFA/I pili from ETEC E. coli (PDB ID number: 3F83) <http://www.rcsb.org/pdb/explore.do?structureId=3f83> (2018).
- National Center for Biotechnology Information. PubChem Compound Database; CID = 91852492 <https://pubchem.ncbi.nlm.nih.gov/compound/91852492#section=Top> (2018).
- National Center for Biotechnology Information. PubChem Compound Database; CID = 11051152. <https://pubchem.ncbi.nlm.nih.gov/compound/11051152#section=Top> (2018).
- Heggelund, J. E., Varrot, A., Imbert, A. & Krengel, U. Histo-blood group antigens as mediators of infections. *Curr Opin Struct Biol* **44**, 190–200, <https://doi.org/10.1016/j.sbi.2017.04.001> (2017).

29. Coddens, A. *et al.* Recognition of Blood Group ABH Type 1 Determinants by the FedF Adhesin of F18-fimbriated *Escherichia coli*. *J. Biol. Chem* **284**(15), 9713–9726, <https://doi.org/10.1074/jbc.M807866200> (2009).
30. Sjöling, Å., Wiklund, G., Savarino, S. J., Cohen, D. I. & Svennerholm, A. M. Comparative Analyses of Phenotypic and Genotypic Methods for Detection of Enterotoxigenic *Escherichia coli* Toxins and Colonisation Factors. *J. Clin. Microbiol* **45**(10), 3295–3301, <https://doi.org/10.1128/JCM.00471-07> (2007).
31. Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **47**, 1739–1749, <https://doi.org/10.1021/jm0306430> (2004).
32. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242 (2000).
33. Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* **27**, 221–234, <https://doi.org/10.1007/s10822-013-9644-8> (2013).
34. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* **118**, 11225–11236, <https://doi.org/10.1021/ja9621760> (1996).
35. Friesner, R. A. *et al.* Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *Journal of Medicinal Chemistry* **49**, 6177–6196, <https://doi.org/10.1021/jm051256o> (2006).
36. Bell, J. A. *et al.* PrimeX and the Schrödinger computational chemistry suite of programs. *International Tables for Crystallography*. 534–538 (2012).
37. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* **10**, 449–461, <https://doi.org/10.1517/17460441.2015.1032936> (2015).

## Acknowledgements

We gratefully thank Professor Weiping Zhang of Kansas State University, USA for the kind gift of monoclonal anti CfaB and anti CfaE antibodies. We acknowledge the Centre for Cellular Imaging at the Sahlgrenska Academy, University of Gothenburg, Sweden for the use of their imaging equipment, and for the support of their staff. LM, JT and AMS work is supported by the Swedish Research Council (grants 2013-6615 and 2011-3435) and the Swedish Strategic Foundation (grant SB12-0072). JH work is supported by the County Council of Västra Götaland (ALF). SC gratefully acknowledges Professor Leif Eriksson of the Department of Chemistry and Molecular biology, University of Gothenburg, Sweden for his valuable guidance and suggestions in computational analysis of this work. SC acknowledges the Swedish Research Council (grant 2014-3914), the University of Gothenburg Center for Antibiotics Resistance Research, and the Faculty of Science for financial support. The Swedish National Infrastructure for Computing (Grant number 2016/34-30) is gratefully acknowledged for grants of computing time at the Chalmers supercomputing center (C3SE).

## Author Contributions

L.M. performed the infection experiments, analysed all data and wrote the manuscript. J.L. glyco-engineered the CHO-cell lines to express Le<sup>a</sup> and Le<sup>b</sup> determinants. S.C. performed the molecular docking experiments. J.T. constructed the recombinant bacterial strains. A.M.S. and J.H. oversaw the project. All authors reviewed the manuscript before submission.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-29258-0>.

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

**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) 2018