Binding of *Clostridium difficile* Toxins to Human Milk Oligosaccharides

Amr El-Hawiet,[1] Elena N. Kitova,[1] Pavel Kitov,[1] Luiz Eugenio,[2] Kenneth K.S. Ng,[2] George L. Mulvey[3], Tanis C. Dingle[3], Adam Szpacenko,[1] Glen D. Armstrong[3] and John S. Klassen[1]*

1. Department of Chemistry, University of Alberta, Edmonton, AB

2. Department of Biological Sciences, University of Calgary, Calgary, AB

3. Department of Microbiology & Infectious Diseases, University of Calgary, Calgary, AB

[†]Alberta Ingenuity Centre for Carbohydrate Science

*Corresponding Author's address:

Department of Chemistry

University of Alberta

Edmonton, AB CANADA T6G 2G2

Email: john.klassen@ualberta.ca

Telephone: (780) 492 3501

Fax: (780) 492 8231

Keywords: Clostridium difficile / toxin / human milk oligosaccharide / ligand binding / docking / cytotoxicity / mass spectrometry

Abstract

The binding of recombinant fragments of the C-terminal cell-binding domains of the two large exotoxins, toxin A (TcdA) and toxin B (TcdB), expressed by *Clostridium difficile* and a library consisting of the most abundant neutral and acidic human milk oligosaccharides (HMOs) was examined quantitatively at 25 °C and pH 7 using the direct electrospray ionization mass spectrometry (ES-MS) assay. The results of the ES-MS measurements indicate that both toxins fragments investigated, TcdB-B1 and TcdA-A2, which possess one and two carbohydrate binding sites, respectively, bind specifically to HMOs ranging in size from tri- to heptasaccharides. Notably, five of the HMOs tested bind to both toxins: Fuc(α1-2)Gal(β1-4)Glc, Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)Glc, Fuc(α1-2)Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)Glc, Gal(β1-3)[Fuc(α1-4)]GlcNAc(β1-3)Gal(β1-4)Glc and Gal(β1-4)[Fuc(α1-3)]GlcNAc(β1-3)Gal(β1-4)Glc. However, the binding of the HMOs is uniformly weak, with apparent affinities $\leq 10^3$ M^{-1}. The results of molecular docking simulations, taken together with the experimental binding data, suggest that a disaccharide moiety (lactose or lactosamine) represents the core HMO recognition element for both toxin fragments. The results of a Verocytotoxicity neutralization assay reveal that the HMOs do not significantly inhibit the cytotoxic effects of TcdA or TcdB. The absence of protection is attributed to the very weak intrinsic affinities that the toxins exhibit towards the HMOs.

Introduction

Clostridium difficile is a Gram-positive, spore forming, strict anaerobic bacterium responsible for a variety of toxin-mediated gastrointestinal diseases that range in severity from antibiotic-associated diarrhea to pseudomembranous colitis (Bartlett 2008; Kelly and LaMont 1998). The emergence of a new and more virulent strain in North America and Europe has been linked to increased morbidity and mortality. Although *C. difficile* can produce up to six different toxins, the main virulence factors are the two exotoxins, toxin A (TcdA) and toxin B (TcdB) (Curry et al. 2007; Voth and Ballard 2005). These two toxins, which share 47% amino acid sequence identity, belong to the large clostridial glucosylating toxin family (Voth and Ballard 2005). Both toxins catalyze the transfer of glucose onto the Rho family of GTPases, leading to a disruption of the cytoskeleton and cell death (Lyras et al. 2009). Although both TcdA and TcdB share a common glucosyltransferase activity and overall structure, differences in structure, substrate specificity and receptor binding appear to contribute towards different cytotoxic mechanisms (Voth and Ballard 2005; Lyras et al. 2009; Ballard 2010; Kuehne et al, 2010; Pruitt et al. 2010). The two toxins appear to exert complementary effects to synergistically disrupt the intestinal epithelium during pathogenesis.

Like all members of the clostridial toxin group, TcdA and TcdB are large (308 and 250 kDa, respectively) single-subunit polypeptides, whose structures appear to be organized into three regions: (i) an N-terminal glucosyltransferase domain; (ii) a central region containing cysteine and aspartyl proteolytic activities, as well as a hydrophobic region, which is important for translocating the toxins across the cell membrane; and (iii) a highly repetitive C-terminal region, which appears to be primarily responsible for

receptor binding (Dove et al. 1990; Voneichelstreiber et al. 1992; Voneichelstreiber et al. 1996; Rupnik et al. 2009; Pruitt et al. 2010). Although *C. difficile* infections can usually be controlled by treatment with broad-spectrum antibiotics, like metronidazole and vancomycin, existing therapeutic approaches are not effective for treating novel hypervirulent and drug-resistant strains, as well as many cases of relapse or reinfection due to the continued disruption of normal bacterial flora following antibiotic treatment (Gerding and Johnson 2010). As a result, alternative therapeutic strategies are required to prevent *C. difficile* from colonizing the intestinal tract and to neutralize the cytotoxic effects of TcdA and TcdB. It has been proposed that such a therapy may involve host cell receptor analogs in various forms that are able to competitively inhibit TcdA and TcdB from binding to the surface of human intestinal epithelial cells (Heerze et al. 1994). The rationale behind this approach is to provide toxins with decoy ligands in the gastrointestinal tract that will divert them from their native receptors on the host cell surface, thus sequestering the toxins and facilitating their elimination from the body.

The specific functional receptors for TcdA and TcdB toxins in humans have yet to be positively identified (Dallas and Rolfe 1998; Krivan et al. 1986; Rolfe and Song 1995). Currently, the only known native receptor for TcdA is the trisaccharide αGal(1,3)βGal(1,4)βGlcNAc), which is found on the surface of rabbit erythrocytes, hamster brush border membranes, bovine thyroglobulin and both Immunoglobulin (Ig) and non-Ig components of human milk (Dallas and Rolfe 1998; Krivan et al. 1986; Rolfe and Song 1995). Recently, it was shown that the related trisaccharide αGal(1,3)βGal(1,4)βGlc and its analogs bind specifically, albeit weakly, to fragments of TcdA and TcdB (Dingle et al. 2008). TcdA also binds to Lewis X, Y, and I glycan

sequences, which are expressed on the surface of human intestinal epithelial cells (Smith et al. 1997; Tucker and Wilkins 1991).

Human milk oligosaccharides (HMOs) are known to protect newborns from variety of infectious diseases (Newburg 2009). For example, HMOs have been shown to inhibit the attachment of *Streptococcus pneumoniae*, *Norwalk-like* virus, *Haemophilus influenzae* to host cells (Andersson et al. 1986; Newburg et al. 2005). Furthermore, fucosylated oligosaccharides from human milk have been found to protect infants from the heat stable toxin of *E. coli* and to prevent the binding of *Campylobacter jejuni* to its receptor in human epithelial cells (Cravioto et al. 1991; Ruiz-Palacios et al. 2003). Additionally, the trisaccharide 3′-sialyllactose inhibits the binding of *Helicobacter pylori* to the gastrointestinal epithelium and shows protection against cholera toxin-induced diarrhea (Idota et al. 1995; Martin-Sosa et al. 2002; Ninonuevo et al. 2006). While protection may be due to the prebiotic characteristics of HMOs, it is believed to result primarily from inhibition of binding of pathogens to host cells due to the similarity of HMOs to epithelial cell surface carbohydrates.

In the present study, the potential of HMOs as inhibitors of *C. difficile* TcdA and TcdB was explored. The binding of twenty-one HMOs, which represent the most abundant oligosaccharides in human milk, to fragments of TcdA and TcdB was investigated using the direct electrospray ionization mass spectrometry (ES-MS) assay. Molecular docking simulations were carried out to elucidate the molecular basis of HMO recognition by the toxin fragments. Cytotoxicity neutralization assays were also performed to investigate the inhibitory potential of HMOs on TcdA and TcdB.

Experimental

Proteins

The TcdA-A2 subfragment (A2, MW 29590 Da) and TcdB-B1 subfragment (B1, MW 14860 Da) was expressed in *Escherichia coli* and purified as described previously (Greco et al. 2006; Ho et al. 2005). Purified samples of TcdA-A2 and TcdB-B1 solutions were stored at -80 ^{o}C. Lysozyme (Lyz, MW 14310 Da) and α-lactalbumin (LA, MW 14200 Da), which served as reference proteins (P_{ref}) for the ES-MS binding assays, were purchased from Sigma–Aldrich Canada (Oakville, ON) and used without further purification.

Carbohydrates

The HMOs library consisted of lactose (Gal(β1-4)Glc), MW 342 Da, (**L1)**; 2′-fucosyl-lactose, Fuc(α1-2)Gal(β1-4)Glc, MW 488 Da, 2FL (**L2**); 3-fucosyl-lactose, Gal(β1-4)[Fuc(α1-3)]Glc, MW 488 Da, 3FL (**L3**); difucosyl-lactose, Fuc(α1-2)Gal(β1-4)[Fuc(α1-3)]Glc, MW 634 Da, LDFT (**L4**); 3′-sialyl-lactose, Neu5Ac(α2-3)Gal(β1-4)Glc, MW 633 Da, 3SL (**L5**); 6′-sialyl-lactose, Neu5Ac(α2-6)Gal(β1-4)Glc, MW 633 Da, 6SL (**L6**); 3′-sialyl-3-fucosyl-lactose, Neu5Ac(α2-3)Gal(β1-4)[Fuc(α1-3)]Glc, MW 779 Da, 3′NeuAc-3FL (**L7**); lacto-N-tetraose, Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)Glc, MW 708 Da, LNT (**L8**); lacto-N-fucopentaose I, Fuc(α1-2)Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)Glc, MW 853 Da, LNFI (**L9**); lacto-N-fucopentaose II, Gal(β1-3)[Fuc(α1-4)]GlcNAc(β1-3)Gal(β1-4)Glc, MW 853 Da, LNFII (**L10**); lacto-N-difucohexaose I, Fuc(α1-2)Gal(β1-3)[Fuc(α1-4)]GlcNAc(β1-3)Gal(β1-4)Glc, MW 1000 Da, LNDI (**L11**); difucosyl-lacto-N-hexaose(a), Gal(β1-4)[Fuc(α1-3)]GlcNAc(β1-6)[Fuc(α1-2)Gal(β1-3)GlcNAc(β1-3)]Gal(β1-4)Glc, MW 1365 Da, F2-LNH a (**L12**); sialyl-lacto-N-tetraose a, Neu5Ac(α2-3)Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)Glc, MW 998 Da, LST a (**L13**);

sialyl-lacto-N-tetraose b, Neu5Ac(α2-6)[Gal(β1-3)]GlcNAc(β1-3)Gal(β1-4)Glc, , MW 998 Da, LST b (**L14**); disialyl-lacto-N-tetraose, Neu5Ac(α2-3)Gal(β1-3)[Neu5Ac(α2-6)]GlcNAc(β1-3)Gal(β1-4)Glc, MW 1290 Da, disialyl-LNT (**L15**); sialyl-fucosyl-lacto-N-tetraose, Neu5Ac(α2-3)Gal(β1-3)[Fuc(α1-4)]GlcNAc(β1-3)Gal(β1-4)Glc, MW 1145 Da, sialyl-Le^a or Sia-LNF III (**L16**); sialyl-lacto N-fucopentaose V, Fuc(α1-2)Gal(β1-3)[Neu5Ac(α2-6)]GlcNAc(β1-3)Gal(β1-4)Glc, MW 1145 Da, Sia-LNF V (**L17**); lacto-N-neo-tetraose, Gal(β1-4)GlcNAc(β1-3)Gal(β1-4)Glc, MW 708 Da, LnNT (**L18**); lacto-N-fucopentaose III, Gal(β1-4)[Fuc(α1-3)]GlcNAc(β1-3)Gal(β1-4)Glc, MW 853 Da, LNFIII (**L19**); difucosyl para-lacto-N-hexaose, Gal(β1-3)[Fuc(α1-4)]GlcNAc(β1-3)Gal(β1-4)[Fuc(α1-3)]GlcNAc(β1-3)]Gal(β1-4)Glc, MW 1365 Da, Lea/Lex (**L20**); sialyl-lacto-N-tetraose c, Neu5Ac(α2-6)Gal(β1-4)GlcNAc(β1-3)Gal(β1-4)Glc, MW 998 Da, LST c (**L21)**. The HMOs were purchased from Sigma–Aldrich (Oakville, ON) (**L1**) and IsoSep AB (Sweden) (**L2-L21**). Stock solutions were prepared by dissolving all ligands with ultrafiltered water (Milli-Q, Millipore) at a concentration of 1 mM and stored at -20 ^{o}C until used.

Isolation of HMOs from human milk

Human milk donations were obtained from the Alberta Children's Hospital, Calgary, Alberta, Canada. Oligosaccharides were extracted as previously described (Ward et al. 2006). Raw milk (1 L) was centrifuged at 5,000 ×g for 30 min at 4°C, and the fat was removed. Ethanol (2 L) was added, and the solution was incubated overnight at 24°C. The precipitate was removed by centrifugation at 5,000 ×g for 30 min at 4°C, and the solvent was removed by rotary evaporation. The carbohydrate fraction was dissolved in 5 mL of water and the solution was passed through a Bio Gel P-2 (Extra fine,

<45 μm; Bio Rad Laboratories, Hercules, CA, USA) column (2.6 X 100 cm). Elution was performed with 100 mM ammonium acetate at a flow rate of 26 mL/h, and the elution profile was recorded with a refractive Index detector (Waters, differential refractometer R401). A total of 6 (I-VI) HMO fractions (7.5 mL volume each) were collected and freeze-dried. Representative ES mass spectra acquired for aqueous solutions of 1mg/100μL of each HMO fractions are given in Figure S1. Mass and composition of the oligosaccharides detected in the isolated HMO fractions are listed in Table S1 (Supplementary Data).

Direct ES-MS assay

Apparent association constants ($K_{a,app}$) for the fragments TcdA-A2 and TcdB-B1 binding to the library of twenty one HMOs (**L1 – L21)** were evaluated using the ES-MS assay. Complete details of the experimental methodology and data analysis are described elsewhere (Sun et al. 2006; Wang et al. 2003) and only a brief overview is given here. The ES-MS measurements were carried out using a 9.4T Apex II Fourier-transform ion cyclotron resonance (FTICR) MS (Bruker-Daltonics, Billerica, MA). Prior to analysis, the TcdA-A2 and TcdB-B1 solutions were diluted with 50 mM ammonium acetate (pH 7.2) and concentrated using Amicon Ultra-4 centrifugal filters with a molecular weight cut-off of 10,000 Da (Millipore). The concentrations of the TcdA-A2 and TcdB-B1 solutions were measured by UV absorption. Each ES solution was prepared from stock solutions of protein (TcdA-A2 and TcdB-B1) and one of carbohydrate ligands (**L1 – L21)**. Lysozyme (Lyz) and α-lactalbumin (La) were used as reference proteins (P_{ref}) to distinguish specific from nonspecific ligand binding with TcdA-A2 and TcdB-B1, respectively, during the ES-MS measurements.

Docking simulations

Automated molecular docking simulations were conducted with AutoDock Vina 1.1.1 (Trott and Olsen 2009) in conjunction with the MGL Tools 1.5.4 graphical interface (Scripps Research Institute, La Jolla, CA). The crystal structure of the complex between TcdA-A2 and αGal(1,3)βGal(1,4)βGlcNAc (Greco et al. 2006) (Protein Databank entry 2G7C) was used without modifications except for the addition of polar hydrogen atoms (hydrogens bound to hetero atoms) as required by AutoDock. The original bound ligand and all water molecules were removed, followed by addition of polar hydrogens to the crystal structure using AutoDock Tools. For TcdB-B1, a homology model was generated by Modeller (Fiser and Sali 2003) with template-based refinement of the binding site as described below. Ligand structures were built using the program Insight II 2005 (Accelrys Inc.) and their energies were minimized using the standard AMBER force field with Homans' parameters for carbohydrates (Homans 1990). The dielectric constant was set at 4 distance dependent, 1-4 parameters were scaled by 0.5. The Grid box was centered on the ring oxygen of the β-galactose moiety of the original oligosaccharide ligand found in the crystal structure while box parameters were set at 30 Å in any dimension. The proteins were regarded as rigid, while all non-ring bonds in ligands were set as active (flexible). Alternatively, some or all Φ dihedral angles of inter-glycosidic anomeric bonds were set as inactive (rigid) in order to facilitate finding binding modes in agreement with exo-anomeric effect. The energy range for the docked poses was set at 2.75 kcal mol^{-1} and the number of poses was set at 20. All other docking parameters were set to their default values. When a single docking experiment did not result in a pose in which all the Φ dihedral angles were consistent with the exo-anomeric effect, docking

were repeated several times with different random seeds. These poses, in which all Φ dihedral angles were consistent with exo-anomeric effect were retained, were then further refined by performing geometry optimization (with the protein structure fixed) using the AMBER force field.

Vero cytotoxicity neutralization assays

Each HMO fraction (I-VI) was 3 fold serial diluted in phosphate-buffered (pH 7.2) physiological saline (PBS). The HMO dilutions were then admixed with purified TcdA or TcdB holotoxins (Heerze et al. 1994) diluted in PBS to their CD_{100} concentration; the minimum concentration resulting in a 100% cytopathic effect in the Verocytotoxicity assay. From each of these samples, 20 μL was transferred to a 96-well microtiter plate containing confluent Vero cell monolayers cultivated in MEM tissue culture growth medium excluding fetal bovine serum (FBS). The final concentration of the HMO fractions in the first well of the Vero cell plate was 2.72 mg/ml. Wells containing TcdA- or TcdB-specific polyclonal rabbit antisera serial diluted in PBS admixed with TcdA or TcdB served as positive inhibition controls and wells containing PBS alone served as the negative inhibition controls. The microtiter plates were then incubated for 4 h at 37°C before the medium in each well was removed and replaced with fresh MEM supplemented with 10% FBS. The plates were incubated for an additional 48 h and cell viability was subsequently assessed by the conventional Giemsa staining technique and the results were recorded using a microtitre plate reader set to an absorbance of 630 nm.

Results and discussion

ES-MS binding measurements

The direct ES-MS assay was used to test for specific binding between the A2 fragment of TcdA and the B1 fragment of TcdB and each of the twenty-one HMOs (**L1**-**L21**) and to quantify their affinities at pH 7 and 25 °C. A detailed description of the ES-MS results obtained for **L8**, which binds to both toxin fragments and served as a model ligand for establishing appropriate experimental and instrumental conditions for the binding measurements, is given below followed by a summary of the results obtained for the other HMOs.

Shown in Figures 1a and 1b are representative ES mass spectra acquired for solutions of aqueous ammonium acetate (10 mM), A2 (75 μM), Lyz (12 μM) and **L8** at 50 μM and 100 μM, respectively. As noted above, Lyz served as P_{ref} for the binding measurements performed on A2. Inspection of the ES mass spectra reveals signals corresponding to protonated ions of free (unbound) A2 fragment, as well as the A2 bound to one or two molecules of **L8**, *i.e.*, $(A2 + q\mathbf{L8})^{n+}$ where $q = 0 – 2$ and n = 9 - 12. Ions corresponding to unbound and bound P_{ref} ions were also detected, *i.e.*, $(P_{ref} + q\mathbf{L8})^{n+}$ where $q = 0 - 2$ and n = 7 - 9, indicating that nonspecific binding of **L8** to A2 occurred during the ES process and contributed to the mass spectrum. The distributions of **L8** molecules bound to A2 and to Lyz determined from the mass spectra (Figures 1a and 1b) are shown in Figures 1c and 1d. Also shown are the distributions of **L8** bound to A2 following correction for nonspecific binding. It can be seen that, under these solution conditions, A2 binds a maximum of one molecule of **L8**. Notably, control experiments, which involved varying the ion source conditions, confirmed that the measured distributions of bound **L8** (after correction for nonspecific binding) were not influenced by in-source (gas phase) dissociation (Wang et al. 2003). Measurements were also

performed using higher concentrations of **L8**, up to 200 μM. However, no ions corresponding to specific (A2 + 2(**L8**)) complex were detected. The average $K_{a,app}$ obtained from these measurements, after correction for nonspecific binding, is 1500 ± 500 M^{-1}.

The absence of signal corresponding to the (A2 + 2(**L8**)) complex appears, at first glance, to be at odds with the x-ray crystal structure of A2 bound to a synthetic derivative of the natural carbohydrate receptor, which demonstrated that the A2 fragment has two equivalent binding sites for the αGal(1,3)βGal(1,4)βGlcNAc trisaccharide, (Greco et al. 2006). However, based on the measured $K_{a,app}$ and an initial A2 concentration of 75 μM, a **L8** concentration in excess of 300 μM would be required to produce a detectable concentration of the (A2 + 2(**L8**)) complex. It was found that **L8** concentrations >200 μM led to significant protein signal suppression and extensive nonspecific ligand binding. As a result, the ES-MS binding measurements were restricted to **L8** concentrations ≤ 200 μM.

Shown in Figures 2a and 2b are ES mass spectra acquired for aqueous solutions of ammonium acetate (10 μM), B1 (20 μM), LA (8 μM) and **L8** at 50 μM and 150 μM, respectively. It should be noted that LA was used as P_{ref} for these measurements because the addition of Lyz (the P_{ref} used with A2) causes precipitation of B1. Protonated ions corresponding to the free and ligand-bound B1 were detected, *i.e.*, $(B1 + q\mathbf{L8})^{n+}$ where q = 0 - 2 and n = 5 and 6. Free and bound P_{ref} ions were also detected, *i.e.*, $(P_{ref} + q\mathbf{L8})^{n+}$ where q = 0 – 2 and n = 6 and 7, indicating that nonspecific binding of **L8** to B1 during the ES process contributed to the mass spectrum. Shown in Figures 2c and 2d are the distributions of **L8** bound to B1 obtained from the mass spectra shown in Figure 2a and

2b, respectively, before and after correction for nonspecific binding. According to these results, and those obtained at different ligand concentrations, the B1 fragment binds a single molecule of **L8**, which is consistent with the presence of a single carbohydrate binding site. The average $K_{a,app}$ value is $1000 \pm 500\ M^{-1}$.

Listed in Table 1 are the results of the ES-MS binding measurements performed on the twenty-one HMOs. Where binding was detected, the affinities were determined based on at least six measurements performed at multiple ligand concentrations. The errors were reported as the pooled standard deviation. In all cases, the ES mass spectra were corrected for nonspecific binding using the reference protein method (Sun et al. 2006). It should be noted that the reported $K_{a,app}$ values depend on both the intrinsic affinity of each binding site and the number of available binding sites. Because the B1 fragment has a single binding site, the apparent and intrinsic affinities are equivalent. However, A2 possesses two equivalent carbohydrate binding sites (Dingle et al. 2008). Therefore, the $K_{a,app}$ values are two times larger than the intrinsic affinity.

The results of the ES-MS measurements indicate that both the A2 and B1 fragments bind specifically to a number of the HMOs investigated, which range in size from tri- to heptasaccharides. Of the twenty-one HMOs tested, A2 exhibits a measurable affinity for eight of them - five neutral (**L2**, **L8**, **L9**, **L10**, **L19**) and three acidic HMOs (**L7, L14**, **L15)** - while B1 binds to eleven of the HMOs - all of the neutral HMOs, except **L3** and **L20**, and two acidic HMOs (**L16** and **L21**). Neither fragment exhibits a measurable affinity for lactose. However, the trisaccharide **L2**, which contains an additional fucose at the non-reducing end of lactose, binds to both fragments. Interestingly, five of the neutral HMOs (**L2**, **L8**, **L9**, **L10**, and **L19**) are recognized by both A2 and B1. This result points

to a degree of structural similarity in the ligand binding sites of the two toxins and raises the possibility of there being common natural human receptors that are recognized by both toxins.

The binding of the toxin fragments to the HMOs is uniformly weak, $K_{a,app} \leq 3100\ M^{-1}$, at pH 7 and 25 ^{o}C. The highest affinity ligand for A2 is **L2,** with an $K_{a,app}$ of 2000 M^{-1}, while the most active ligand for B1 is **L9,** with a $K_{a,app}$ of 3100 M^{-1}. However, it is interesting to note that a number of HMOs bind to the toxins with a much higher affinity than the only known natural receptor, αGal(1,3)βGal(1,4)βGlc, for which binding constants for both A2 and B1 were found to be ~500 M^{-1} (Dingle et al. 2008). The results of this study also indicate that increased complexity of the HMOs does not necessarily result in a significant increase in affinity.

The patterns of monovalent, solution-phase ligand binding observed by ES-MS for A2 show some similarities with the patterns of multivalent, solid-phase ligand binding observed by glycan array screening with the A2 fragment, as well as native TcdA holotoxin (Consortium for Functional Glycomics (CFG), http:// www.functionalglycomics.org). Most importantly, the Le^X trisaccharide, which is identical to **L2** except for the 2-acetamido group in the residue at the reducing end, is one of the tightest-binding ligands in the glycan array screen for A2 and TcdA holotoxin. Moreover, several more complex oligosaccharides containing Le^X at the reducing end are also some of the tightest binding ligands. The results from the glycan array screening confirm a wide range of modifications can be added at the reducing end of the lactose/LacNAc or Le^X core structures, but most of these modifications add little to increase binding affinity. One of the most interesting differences between the ES-MS

method and glycan array screening, which was also seen in an earlier study (Dingle et al. 2008), is the ability of the ES-MS method to measure solution-phase binding constants for ligands of TcdB that have no detectable affinity using the multivalent presentation format of the glycan array. Although the molecular basis for lower affinity multivalent binding in TcdB is not clearly understood at present, differences in the apparent binding affinities of ligands in solution versus multivalent binding interactions with solid-phase glycan arrays have been seen in other systems (Oyelaran and Gildersleeve 2009). As discussed in the earlier ES-MS study, for example, the three-dimensional arrangement of carbohydrate-binding sites in TcdB may not be compatible with the presentation of ligands in the glycan array, a complication that does not affect single-site measurements peformed using the ES-MS method.

Docking analysis

To gain more insight into the interaction modalities of HMOs with *C. difficile* toxin fragments A2 and B1, molecular docking experiments were performed using Autodock Vina. A standard deviation of about 2-3 kcal mol^{-1} in free energy prediction is typical for current computational docking techniques, including AutoDock Vina (Huey et al. 2006; Trott and Olsen 2009). Additionally, although severe intra-molecular steric clashes are avoided during docking in Vina, the internal (conformational) free energy is not taken into account during the final ranking of poses. This can lead to significant errors in ranking of the ligand binding conformations (poses), especially when screening flexible molecules like oligosaccharides. In other words, the lowest energy (top ranked) pose is not always consistent with the structure identified experimentally. Several aspects of the docked conformation may be used to filter docking results, such as the presence of

key amino acids contacts, structural similarities to known ligands, or the availability of unpaired hydrogen bond donors or acceptors in the protein-ligand complex. Except in rare cases, the conformations of carbohydrate ligands co-crystallized with proteins agree with the exo-anomeric effect of oligosaccaharides. The exo-anomeric effect arises from the overlap of a lone electron pair of the exo-anomeric oxygen with σ* of the O_{ring}-C_1 bond. This favorable stereo-electronic interaction stabilizes conformations, in which lone pair and σ* are properly aligned. Although Autodock Vina has no provisions for the exo-anomeric or other stereo-electronic effects in its scoring function in the free energy evaluations, these conformational constraints can be taken into account either by fixing the Φ dihedral angle (O_{ring}-C_1-O_{exo}-C_i) during molecular docking or by filtering out the putative binding modes that fail to satisfy this rule.

To test the reliability of AutoDock Vina for modeling protein-oligosaccharide interactions, the rabbit receptor of TcdA, αGal(1,3)βGal(1,4)βGlcNAc, was prepared using the AMBER force field with Homans' parameters for carbohydrates and docked into the carbohydrate binding site of TcdA-A2 and the results compared to the crystal structure (PDB entry 2G7C). Notably, the selected pose that satisfies the exo-anomeric effect (which was also the lowest energy pose in most of the docking sessions) matches very well (root mean squareue deviation, RMSD <0.1 Å) the pose found in the crystal structure (Figure S2). The impressive accuracy of this molecular docking exercise suggests that AutoDock Vina is a suitable computational tool to probe the binding modalities of flexible oligosaccharide ligands, such as the HMOs, to TcdA-A2.

Using this approach, molecular docking was performed on each of the HMOs that were found by ES-MS to bind to TcdA-A2. Analysis of the docking results reveals that

five of the eight ligands (**L2**, **L7**, **L8**, **L9**, and **L10**) share a common binding motif, GlcNAcβ(1,3)βGal(1,4)βGlc, in which the lactose (Lac) disaccharide fragment represents the core recognition element with the lowest RMSD between these ligands (Figure 3a). In the case of **L19,** a frame shift was identified, wherein it is the LacNAc fragment that binds, and the Lac moiety does not form contacts with the protein (Figure 3b). Notably, this binding motif (involving Lac or LacNAc) matches the position of the LacNAc fragment of the trisaccharide receptor Galα(1,3)βGal(1,4)βGlcNAc identified from the crystal structure (Greco et al. 2006). In the case of **L14** and **L15**, repeated docking experiments failed to identify likely binding poses (i.e., no poses were obtained in which all Φ dihedral angles were consistent with the exo-anomeric effect). This finding may reflect the extreme flexibility of the glycosidic bond in neuraminic acid (Bitzer et al. 2005). The docking results also provide an explanation for the absence of binding in at least two of the HMOs tested. According to the binding poses shown in Figure 3a, the hydroxymethylene group at C6 of β-galactose fits within a distinct indentation in the binding pocket, formed by Glu84, Tyr85, Ile101, Ser121, and Lys122. It is reasonable to expect that any substitutions at this position of Gal will preclude this interaction (or else radically change the mode of binding). The absence of binding observed for **L6** or **L12**, which have sialic acid or Gal(1,4)[Fuc(1,3)]GlcNAc at the C6 position, respectively, can be understood on the basis of this argument.

Molecular docking was also performed to evaluate the interactions between the HMOs and TcdB-B1. Due to the lack of a crystal structure for TcdB-B1, a homology model for TcdB-B1 generated by Modeller (Fiser and Sali 2003) was used. In an effort to "refine" the structure of the binding site, geometry optimization was performed on the B1

fragment in the presence of a virtual ligand (**VL**) (Figure 4a), which was based on the structures of the HMOs that were found to bind both toxin A and toxin B. To construct **VL**, the molecular structures of all the HMOs that bind both TcdA-A2 and TcdB-B1 (**L2**, **L8**, **L9**, **L10** and **L19**) were considered. Since the purpose of the virtual ligand was to induce TcdB-B1 to adopt the bioactive conformation, it was desirable to maximize the number of intermolecular contacts. To this end **L8**, which effectively fills the binding pocket of TcdA-A2, was chosen as the core structure of **VL**. Two branching α-fucose residues were also added, one linked to the GlcNAc in the third position (to resemble **L10**) and the other fucose linked to the reducing end Glc (to resemble **L19**). The Φ and Ψ dihedral angles in the glycosidic bonds of both attached monosaccharides were set equal to those found in **L10** and **L19**. Additionally, the reducing end Glc was replaced with GlcNAc in order to mimic **L19**. The geometry of **VL** was first optimized using the AMBER force field with Homans' parameters for carbohydrates, after which it was introduced to the putative binding site of the homology model for TcdB-B1. The conformation and binding pose of the resulting energy minimized **VL** structure is shown in Figure 4b. Geometry optimization was carried out on the complex of the initial homology model of TcdB-B1 and **VL** with the AMBER force field with Homans' parameters for carbohydrates in a stepwise fashion. First, the protein-ligand interactions were optimized while keeping the protein coordinates fixed; the constraints on protein structure were then released and the energy of the complex was minimized. Comparison of the initial and optimized structures revealed that the most significant structural changes involved the conformations of amino acid side chains that participate in intermolecular contacts; the overall backbone conformations of both protein and oligosaccharide

underwent minimal changes. Also, in the minimized complex, **VL** assumes a conformation in which Φ dihedral angles in all glycosidic bonds are consistent with the exo-anomeric effect.

Docking of eight HMOs (**L2**, **L4**, **L8, L9 L10**, **L18**, **L19** and **L21**) into the refined TcdB-B1 binding site yielded poses that were analogous to the structures found for TcdA-A2 (Figure 5). However, attempts to dock the largest HMOs (**L11**, **L12** and **L16**) failed to produce consistent results, presumably due to the inherent flexibility of these oligosaccharides. Notably, the molecular docking results suggest that, despite a number of differences in amino acid sequence between TcdA and TcdB, the general mode of carbohydrate recognition may be conserved. For example, a lactose disaccharide appears to occupy the central portion of the carbohydrate binding site for both toxins.

It should be noted that the use of a homology model for TcdB in this study likely biases the modeled structure towards the template structure and fails to show some of the true structural differences between TcdA and TcdB. In addition, the docking approach used in this study fails to account for water-mediated interactions, which may play important roles in ligand binding. Nevertheless, analysis of some of the docked structures helps to explain how differences in sequence between TcdA and TcdB, especially the different distributions of negatively and positively charged side chains in the binding pocket, can account for some of the observed differences in ligand specificity. For example, the weak binding of **L7** to TcdA-A2 can be explained, at least in part, as the result of unfavorable electrostatic repulsion between negatively charged carboxyl groups (neuraminic acid and Asp183) and a compensatory favorable electrostatic interaction provided by the proximity of the positively charged guanidino group from Arg193. In

TcdB-B1, this delicate balance appears to be upset by the substitution of the positively charged side chain to the neutral Asn84 and no binding is observed for **L7** (Figure 6).

Cytotoxicity neutralization assay

To investigate the inhibitory potential of HMOs on TcdA and TcdB, Vero cytotoxicity neutralization assays were performed using each of the six fractions (I – VI) of HMOs extracted from human milk samples. The results of the assay reveal that none of the HMO fractions inhibited TcdA or TcdB, while the toxin-specific antisera completely neutralized the cytotoxicity of each holotoxin (Figure 7). These results are not completely unexpected given that TcdA and, likely, TcdB display linear repeats of multiple carbohydrate binding sites in their carboxy terminal cell binding domains (Ho et al. 2005). This arrangement allows the toxin to simultaneously engage multiple glycan receptor sequences on the Vero cell surface, thereby compensating for the observed low affinity interactions between a single glycan receptor sequence and its complimentary carbohydrate binding site. As a consequence, soluble HMOs may not be able to successfully compete with the Vero cells for binding to TcdA or TcdB unless present at very high concentrations. Due to solubility limitations, we were unable to achieve high enough HMO concentrations to demonstrate any possible inhibitory effects.

Conclusions

Using the direct ES-MS assay, the binding of fragments of *C. difficile* toxins TcdA and TcdB with a library HMOs was investigated. The results of the ES-MS measurements indicate that both of the toxin fragments investigated, TcdB-B1 and TcdA-A2, bind specifically to HMOs ranging in size from tri- to heptasaccharides. Notably, five of the HMOs tested bind to both toxins - Fuc(α1-2)Gal(β1-4)Glc, Gal(β1-3)GlcNAc(β1-

3)Gal(β1-4)Glc, Fuc(α1-2)Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)Glc, Gal(β1-3)[Fuc(α1-4)]GlcNAc(β1-3)Gal(β1-4)Glc and Gal(β1-4)[Fuc(α1-3)]GlcNAc(β1-3)Gal(β1-4)Glc, However, the binding of the HMOs is uniformly weak, with apparent affinities $\leq 10^3$ M^{-1}. The results of molecular docking simulations, taken together with the ES-MS binding data, suggest that a disaccharide moiety (lactose or lactosamine) represents the core recognition element for both toxin fragments. Verocytotoxicity neutralization assays indicate that the HMOs do not significantly inhibit the cytotoxic effects of TcdA or TcdB. The absence of protection is attributed to the very weak intrinsic affinities that the toxins exhibit towards the HMOs.

Acknowledgements

The authors would like to thank Alberta Ingenuity Centre for Carbohydrate Science (AICCS), the Natural Sciences and Engineering Research Council (NSERC), and the University of Alberta for supporting this research financially. KN also acknowledges the support of a Senior Scholar award from the Alberta Heritage Foundation for Medical Research.

Abbreviations

ES-MS, electrospray ionization mass spectrometry; FTICR, Fourier-transform ion cyclotron resonance; TcdA, *C.difficile* toxin A; TcdB, *C.difficile* toxin B; $K_{a,app}$, apparent association constant; HMOs, Human milk oligosaccharides; VL, virtual ligand.

References

Andersson B, Porras O, Hanson B, Lagergard T, Svanborgeden C. 1986. Inhibition of attachment of *Streptococcus-Pneumoniae* and *Haemophilus-Influenzae* by human-milk and receptor oligosaccharides. *J. Infect. Dis.* 153: 232-237.

Bartlett JG. 2008. Historical perspectives on studies of *Clostridium difficile* and *Clostridium difficile* infection. *Clin. Infect. Dis*. 46: S4-S11.

Ballard JD. 2010. Medical microbiology: A toxin contest. *Nature* 467(7316):665-666.

Bitzer RS, Barbosa AGH, da Silva CO, Nascimento MAC. 2005.On the generalized valence bond description of the anomeric and exo-anomeric effects: an ab initio conformational study of 2 methoxytetrahydropyran. *Carbohydr. Res.* 340: 2171-2184

Cravioto A, Tello A, Villafan H, Ruiz J, Delvedovo S, Neeser JR. 1991. Inhibition of localized adhesion of Enteropathogenic *Escherichia-Coli* to Hep-2 cells by immunoglobulin and oligosaccharide fractions of human colostrum and breast-milk. *J. Infect. Dis.*163: 1247-1255.

Curry SR, Marsh JW, Muto CA, O'Leary MM, Pasculle AW, Harrison LH. 2007. TcdC genotypes associated of *Clostridium difficile* with severe TcdC truncation in an epidemic clone and other strains of *Clostridium difficile. J. Clin. Microbiol.* 45:215-221.

Dallas SD, Rolfe RD. 1998. Binding of *Clostridium difficile* toxin A to human milk secretory component. *J. Med. Microbiol*. 47: 879-888.

Dingle T, Wee S, Mulvey GL, Greco A, Kitova EN, Sun JX, Lin SJ, Klassen JS, Palcic MM, Ng KKS, Armstrong GD. 2008. Functional properties of the carboxy-terminal host cell-binding domains of the two toxins, TcdA and TcdB, expressed by *Clostridium difficile*. *Glycobiology*. 18: 698-706.

Dove CH, Wang SZ, Price SB, Phelps CJ, Lyerly DM, Wilkins TD, Johnson JL. 1990. Molecular Characterization of the *Clostridium-difficile* Toxin-A Gene. *Infect. Immun*. 58: 480-488.

Fiser A, Sali A. 2003. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol*. 374: 461-491.

Gerding DN, Johnson S. 2010. Healthcare Epidemiology: Management of *Clostridium difficile* Infection: Thinking Inside and Outside the Box. *Clin. Infect. Dis*. 51(11):1306–1313.

Greco A, Ho JGS, Lin SJ, Pacilic MM, Rupnik M, Ng KKS. 2006. Carbohydrate recognition *by Clostridium difficile* toxin A. *Nat. Struct. Mol. Biol.* 13: 460-461.

Heerze LD, Kelm MA, Talbot JA, Armstrong GD. 1994. Oligosaccharide sequences attached to an inert support (Synsorb) as potential therapy for antibiotic-associated diarrhea and *Pseudomembranous Colitis*. *J. Infect. Dis.* 169: 1291-1296.

Homans, SW 1990. A molecular mechanical force-Field for the conformational analysis of oligosaccharides: comparison of theoretical and crystal structrues of Man α1-3Man β1-4 GlcNAc. *Biochemistry*. 29 (39):9110-9118.

Huey R, Morris GM, Olson AJ, Goodsell DS. 2006. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* 28:1145-52.

Ho JGS, Greco A, Rupnik M, Ng KKS. 2005. Crystal structure of receptor-binding C-terminal repeats from *Clostridium difficile* toxin A. *Proc. Natl. Acad. Sci*. 102: 18373-18378.

Idota T, Kawakami H, Murakami Y, Sugawara M. 1995. Inhibition of Cholera-Toxin by human-milk fractions and sialyllactose. *Biosci. Biotechnol. Biochem*. 59: 417-419.

Kelly CP, LaMont JT. 1998. *Clostridium difficile* infection. *Annu. Rev. Med*. 49: 375-390.

Krivan HC, Clark GF, Smith DF, Wilkins TD. 1986. Cell-surface binding-site for *Clostridium-difficile* enterotoxin - evidence for a glycoconjugate containing the sequence Gal-alpha-1-3Gal-beta-1-4GlcNAc. *Infect. Immun*. 53: 573-581.

Kuehne SA, Cartman ST, Heap JT, Kelly ML, Cockayne A, Minton NP. 2010. The role of toxin A and toxin B in Clostridium difficile infection. *Nature* 467(7316):711-713.

Lyras D, O'Connor JR, Howarth PM, Sambol SP, Carter GP, Phumoonna T, Poon R, Adams V, Vedantam G, Johnson S, Gerding DN, Rood JL. 2009. Toxin B is essential for virulence of *Clostridium difficile*. *Nature*. 458: 1176-1181.

Martin-Sosa S, Martin MJ, Hueso P. 2002. The sialylated fraction of milk oligosaccharides is partially responsible for binding to enterotoxigenic and uropathogenic *Escherichia coli* human strains. *J. Nutr.*132: 3067-3072.

Newburg DS, Ruiz-Palacios GM, Morrow AL. 2005. Human milk glycans protect infants against enteric pathogens. *Annu. Rev. Nutr.* 25: 37-58.

Newburg DS. 2009. Neonatal protection by an innate immune system of human milk consisting of oligosaccharides and glycans. *J. Anim Sci*. 87: 26-34.

Ninonuevo MR, Park Y, Yin HF, Zhang JH, Ward RE, Clowers BH, German JB, Freeman SL, Killen K, Grimm R, Lebrilla CB. 2006. A strategy for annotating the human milk glycome. *J. Agric. Food Chem*. 54: 7471-7480.

Oyelaran O, Gildersleeve JC. 2009. Glycan arrays: recent advances and future challenges. *Curr. Opin. in chem. Biol.* 13(4): 406-413.

Pruitt RN, Chambers MG, Ng KK, Ohi MD, Lacy DB. 2010 Structural organization of the functional domains of Clostridium difficile toxins A and B. *Proc. Natl. Acad. Sci. U S A* 107(30):13467-13472.

Rolfe RD, Song W. 1995. Immunoglobulin and nonimmunoglobulin components of human-milk Inhibit *Clostridium-difficile* Toxin A-receptor binding. *J. Med. Microbiol*. 42: 10-19.

Ruiz-Palacios GM, Cervantes LE, Ramos P, Chavez-Munguia B, Newburg DS. 2003. Campylobacter jejuni binds intestinal H(O) antigen (Fuc alpha 1, 2Gal beta 1, 4GlcNAc), and fucosyloligosaccharides of human milk inhibit its binding and infection. *J. Biol. Chem.* 278: 14112-14120.

Rupnik M, Wilcox MH, Gerding DN. 2009. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat. Rev. Microbiol.* 7(7) 526-536.

Smith JA, Cooke DL, Hyde S, Borriello SP, Long RG. 1997. *Clostridium difficile* toxin A binding to human intestinal epithelial cells. *J. Med. Microbiol*. 46: 953-958.

Sun JX, Kitova EN, Wang WJ, Klassen JS. 2006. Method for distinguishing specific from nonspecific protein-ligand complexes in nanoelectrospray ionization mass spectrometry. *Anal. Chem.* 78: 3010-3018.

Trott O, Olson AJ. 2009. Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J.Comput.Chem*. 31: 455-461.

Tucker KD, Wilkins TD. 1991. Toxin-A of *Clostridium-difficile* binds to the human carbohydrate antigens-I, antigens-X, and antigens-Y. *Infect. Immun.*59: 73-78.

Voneichelstreiber C, Sauerborn M, Kuramitsu HK. 1992. Evidence for a modular structure of the homologous repetitive c-terminal carbohydrate-binding sites of *Clostridium-difficile* toxins and *Streptococcus-mutans* Glucosyltransferases. J. Bacteriol.174: 6707-6710.

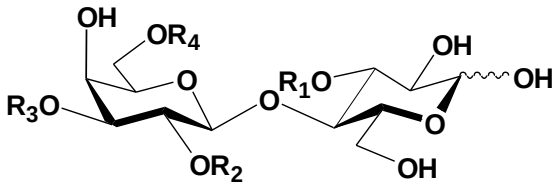
Voneichelstreiber C, Boquet P, Sauerborn M, Thelestam M. 1996. Large clostridial cytotoxins - A family of glycosyltransferases modifying small GTP-binding proteins. *TIM*. 4: 375-382.

Voth DE, Ballard JD. 2005. *Clostridium difficile* toxins: Mechanism of action and role in disease. *Clin. Microbiol. Rev.*18: 247-263.

Wang WJ, Kitova EN, Klassen JS. 2003. Influence of solution and gas phase processes on protein-carbohydrate binding affinities determined by nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* 75: 4945-4955.

Ward RE, Ninonuevo M, Mills DA, Lebrilla CB, German JB. 2006. In vitro fermentation of breast milk oligosaccharides by *Bifidobacterium infantis* and *Lactobacillus gasseri*. *Appl. Environ. Microbiol*. 72: 4497-4499.

Table 1. Apparent association constants, $K_{a,app}$ (units of 10^2 M^{-1}) for binding of the HMOs (**L1** – **21**) with TcdA-A2 and TcdB-B1 fragments, determined at 25 ºC and pH 7 by the direct ES-MS assay. [Rows corresponding to parent compound of each series is shown in bold. Error values correspond to one standard deviation. NB (No Binding) no binding detected.]



HMO	R_1	R_2	R_3	R_4	$K_{a,app}$ (A2)	$K_{a,app}$ (B1)
L1	**H**	**H**	**H**	**H**	**NB**	**NB**
L2	H	Fuc	H	H	20 ± 8	12 ± 5
L3	Fuc	H	H	H	NB	NB
L4	Fuc	Fuc	H	H	NB	10 ± 3
L5	H	H	Neu5Ac	H	NB	NB
L6	H	H	H	Neu5Ac	NB	NB
L7	Fuc	H	Neu5Ac	H	7 ± 3	NB
L8	**H**	**H**	**Gal(β1-3)GlcNAc**	**H**	**15 ± 5**	**10 ± 5**
L9	H	H	Fuc(α1-2)Gal(β1-3)GlcNAc	H	8 ± 1	31 ± 2

L10	H	H	Gal(β1-3)[Fuc(α1-4)]GlcNAc	H	7 ± 2	8 ± 4
L11	H	H	Fuc(α1-2)Gal(β1-3)[Fuc(α1-4)]GlcNAc	H	NB	18 ± 9
L12	H	H	Fuc(α1-2)Gal(β1-3)GlcNAc	Gal(β1-4)[Fuc(α1-3)]GlcNAc	NB	21 ± 5
L13	H	H	Neu5Ac(α2-3)Gal(β1-3)GlcNAc	H	NB	NB
L14	H	H	Neu5Ac(α2-6)[Gal(β1-3)]GlcNAc	H	11 ± 2	NB
L15	H	H	Neu5Ac(α2-3)Gal(β1-3)[NeuAc(α2-6)]GlcNAc	H	7 ± 2	NB
L16	H	H	Neu5Ac(α2-3)Gal(β1-3) [Fuc(α1-4)]GlcNAc	H	NB	11 ± 6
L17	H	H	Fuc(α1-2)Gal(β1-3)[NeuAc(α2-6)]GlcNAc	H	NB	NB
L18	**H**	**H**	**Gal(β1-4)GlcNAc**	**H**	**NB**	**15 ± 2**
L19	H	H	Gal(β1-4) [Fuc(α1-3)]GlcNAc	H	17 ± 2	9 ± 4
L20	H	H	Gal(β1-3)[Fuc(α1-4)]GlcNAc(β1-3)Gal(β1-4)[Fuc(α1-3)]GlcNAc	H	NB	NB
L21	H	H	Neu5Ac(α2-6)Gal(β1-4)GlcNAc	H	NB	20 ± 5

Figure captions

Figure 1. ES mass spectra of aqueous solutions containing (a) 75 μM A2 and 50 μM **L8**, (b) 75 μM A2 and 100 μM **L8**, at pH 7 and 25°C. A P_{ref} (12 μM) was added to each solution to quantify the extent of nonspecific protein-ligand binding during the ES process. (c) and (d) Normalized distributions of **L8** bound to proteins determined from ES mass spectra acquired for the solutions described in (a) and (b), respectively.

Figure 2. ES mass spectra of aqueous solutions containing (a) 20 μM B1 and 50 μM **L8**, (b) 20 μM B1 and 150 μM **L8**, at pH 7 and 25°C. A P_{ref} (8 μM) was added to each solution to quantify the extent of nonspecific protein-ligand binding during the ES process. (c) and (d) Normalized distributions of **L8** bound to proteins determined from ES mass spectra acquired for the solutions described in (a) and (b), respectively.

Figure 3. (a) Representative poses found for docking of **L2**, **L7**, **L8**, **L9** and **L10** and also shown the position of the trisaccharide Galα(1,3)βGal(1,4)βGlcNAc (shown in blue) co-crystallized with TcdA-A2 (PDB entry 2G7C). For clarity, the structures are truncated beyond the trisaccharide GlcNAcβ(1,3) βGal(1,4)βGlc. (b) Frame shift observed in docking of **L19** to TcdA-A2 compared with the reference trisaccharide Galα(1,3)βGal(1,4)βGlcNAc (shown in blue).

Figure 4. (a) Structure of virtual Ligand (**VL**). (b) Structure of **VL** docked in the refined TcdB-B1 binding site.

Figure 5. Preferred poses of oligosaccharides **L2**, **L4**, **L8**, **L9**, **L10, L18**, **L19** and **L21** docked into refined TcdB-B1 binding site. Lactose fragments are colored in blue.

Figure 6. Preferred poses of oligosaccharide **L7** docked into carbohydrate-binding sites of (a) TcdA-A2 and (b) TcdB-B1. Spatially equivalent residues that form a salt bridge in TcdA-A2 (Arg-193 and Asp-183) but fail to form a similar interaction in TcdB (Asn-84 and Asp-65) are labeled. A semi-transparent representation of the solvent-accessible surface of the protein is also drawn.

Figure 7. HMO Verocytotoxicity neutralization assays. Vero cell monolayers were incubated for 4 h at 37^{o}C with 3 fold serial dilutions of HMO fractions I, II, III, IV, V, and VI or TcdA- (a and b) or TcdB-specific (c and d) rabbit polyclonal antisera.

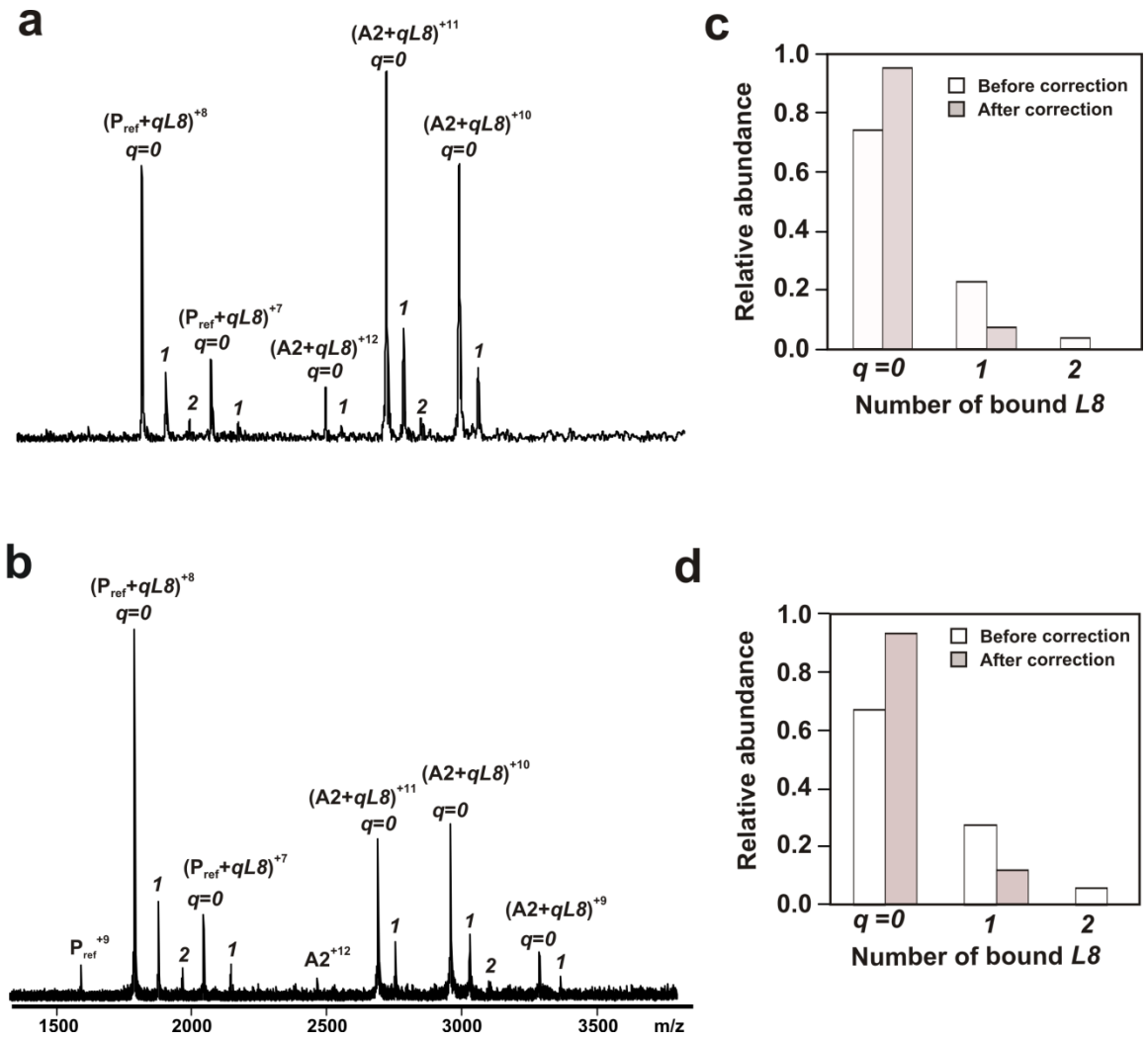

a
(A2+qL8)+11
q=0
(Pref+qL8)+8
q=0
(A2+qL8)+10
q=0
(Pref+qL8)+7
q=0
(A2+qL8)+12
q=0
1
2
1
1
2
1
1
b
(Pref+qL8)+8
q=0
(A2+qL8)+11
q=0
(A2+qL8)+10
q=0
(Pref+qL8)+7
q=0
(A2+qL8)+9
q=0
Pref+9
A2+12
1
2
1
1
1
2
1
1500
2000
2500
3000
3500
m/z
c
Relative abundance
1.0
0.8
0.6
0.4
0.2
0.0
Before correction
After correction
q =0
1
2
Number of bound L8
d
Relative abundance
1.0
0.8
0.6
0.4
0.2
0.0
Before correction
After correction
q =0
1
2
Number of bound L8

Figure 1

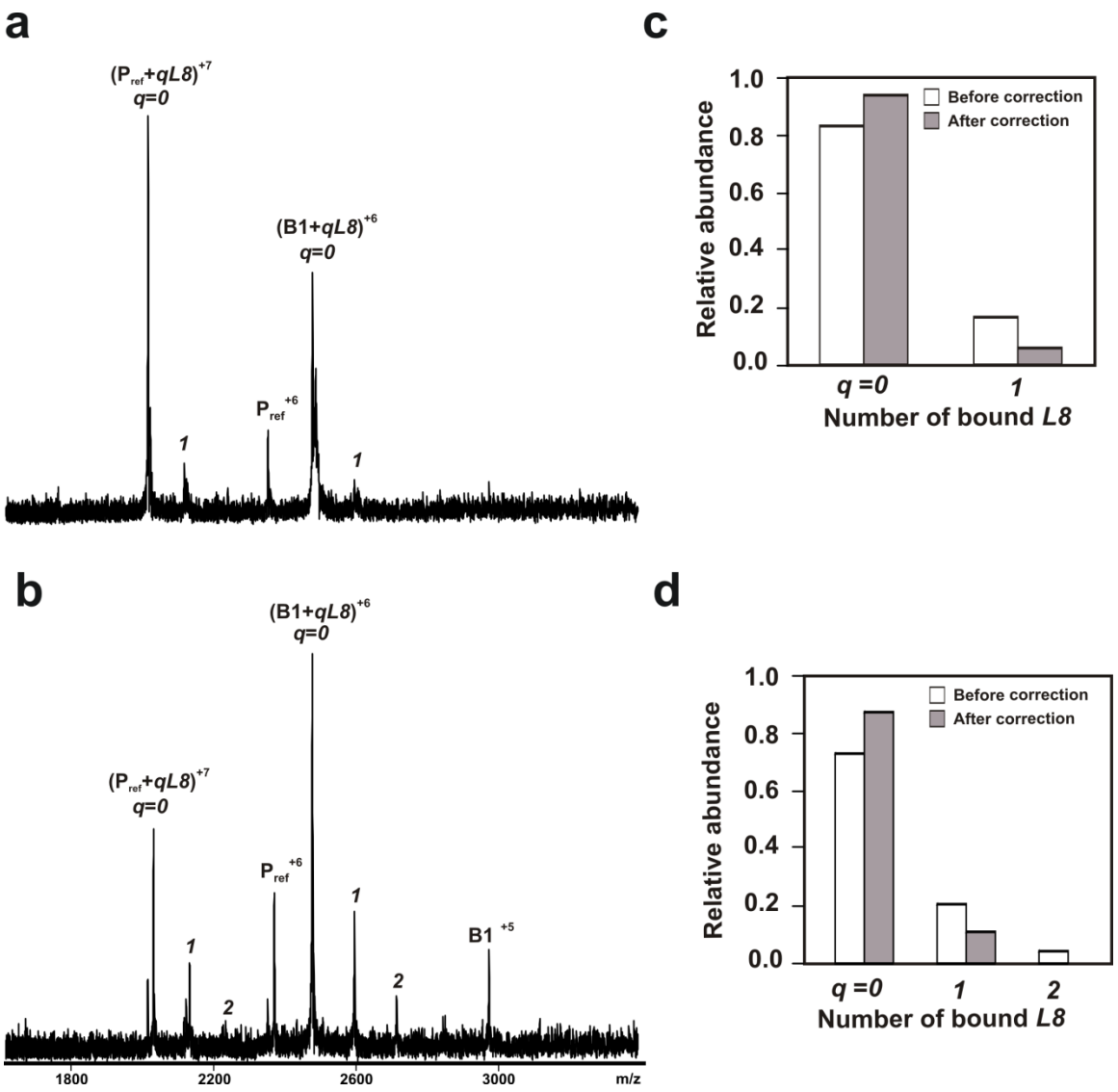
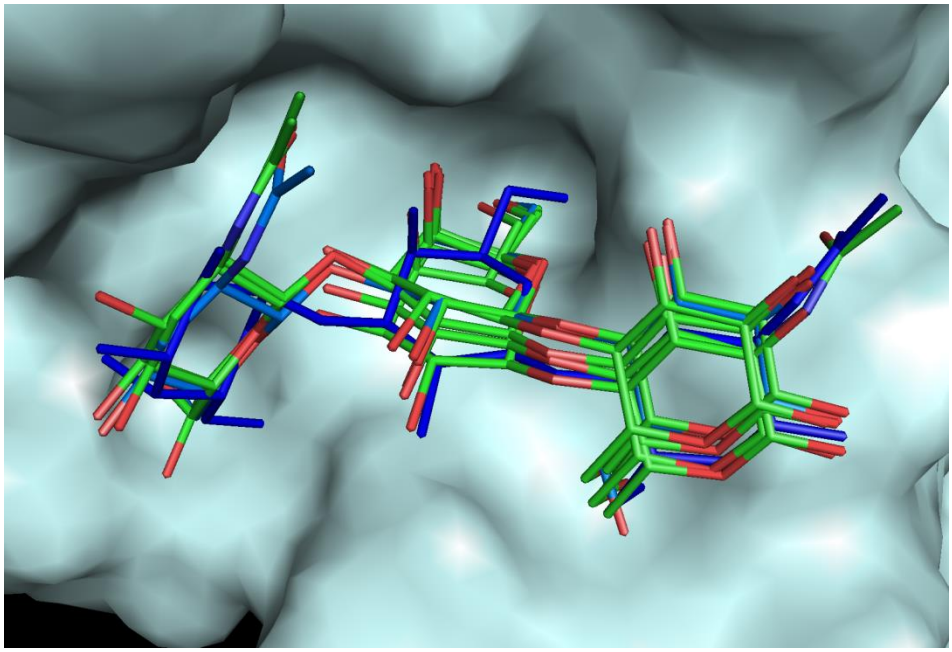

a
(P$_{ref}$+qL8)$^{+7}$
q=0
(B1+qL8)$^{+6}$
q=0
P$_{ref}$$^{+6}$
1
1
b
(B1+qL8)$^{+6}$
q=0
(P$_{ref}$+qL8)$^{+7}$
q=0
P$_{ref}$$^{+6}$
1
1
2
2
B1^{+5}
1800
2200
2600
3000
m/z
c
Relative abundance
1.0
0.8
0.6
0.4
0.2
0.0
Before correction
After correction
q =0
1
Number of bound L8
d
Relative abundance
1.0
0.8
0.6
0.4
0.2
0.0
Before correction
After correction
q =0
1
2
Number of bound L8

Figure 2

a



b

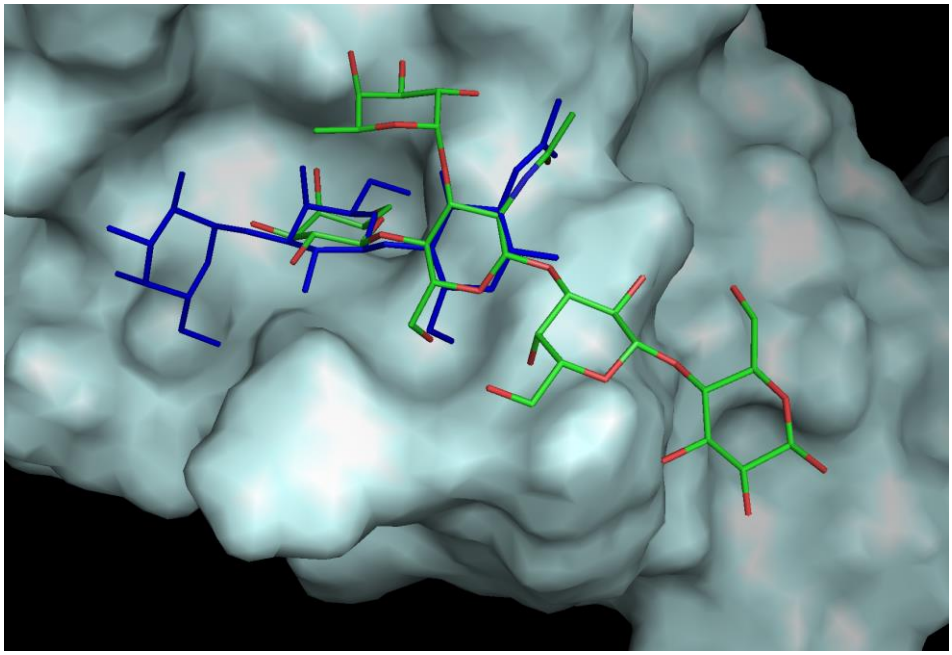
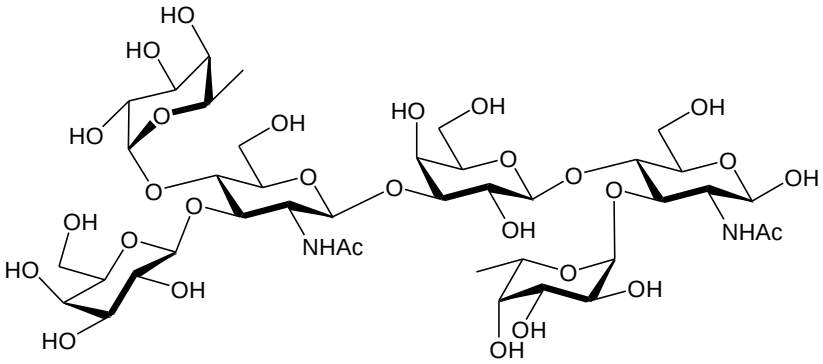


Figure 3

a



VL

b

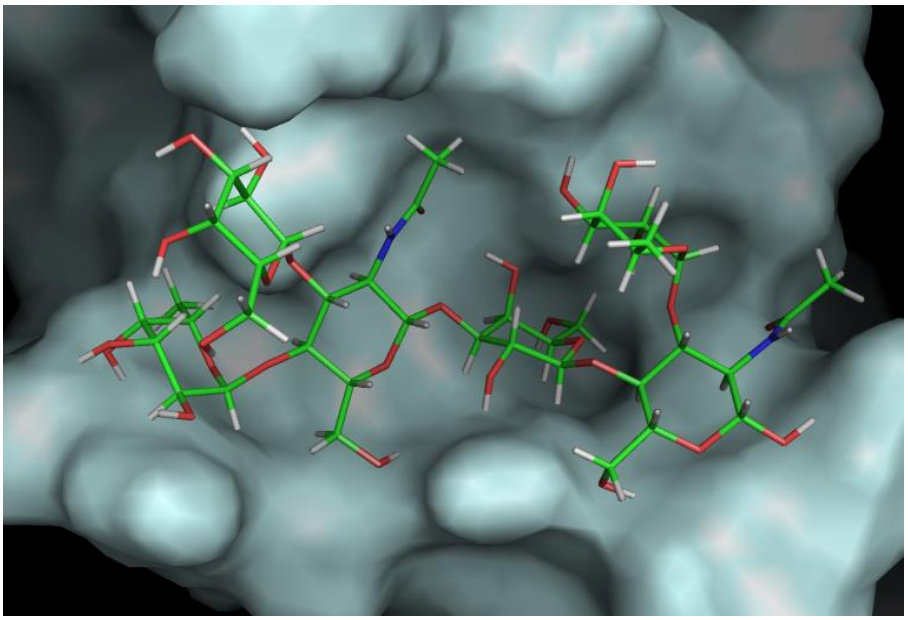


Figure 4

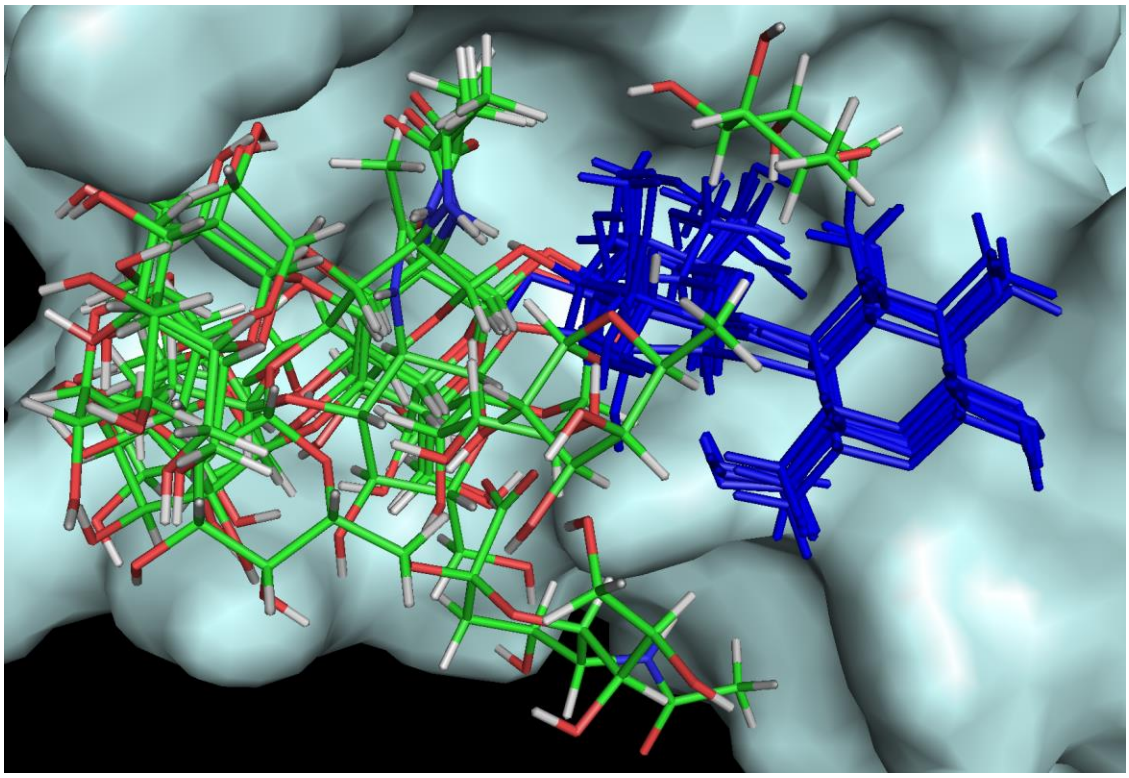


Figure 5

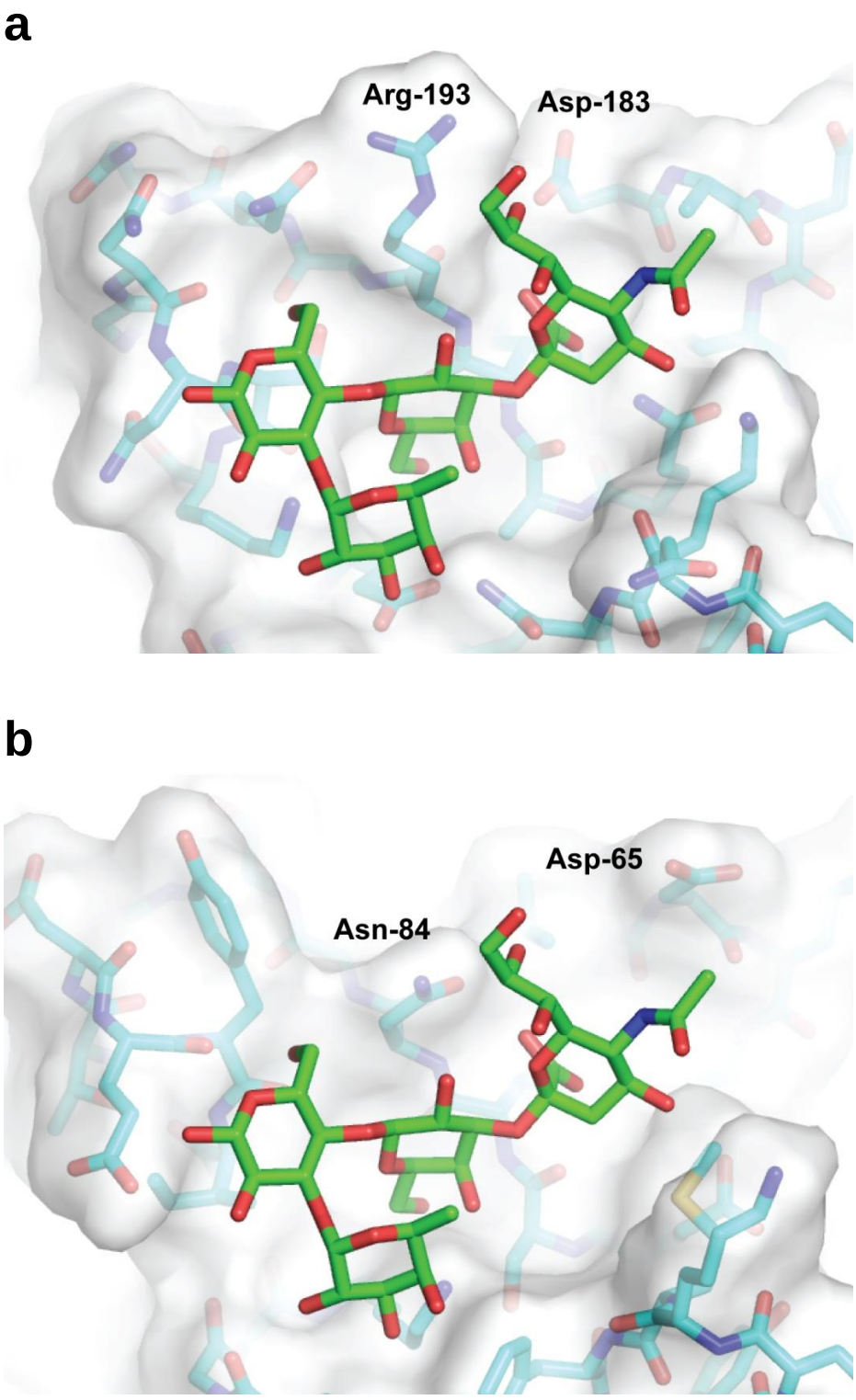
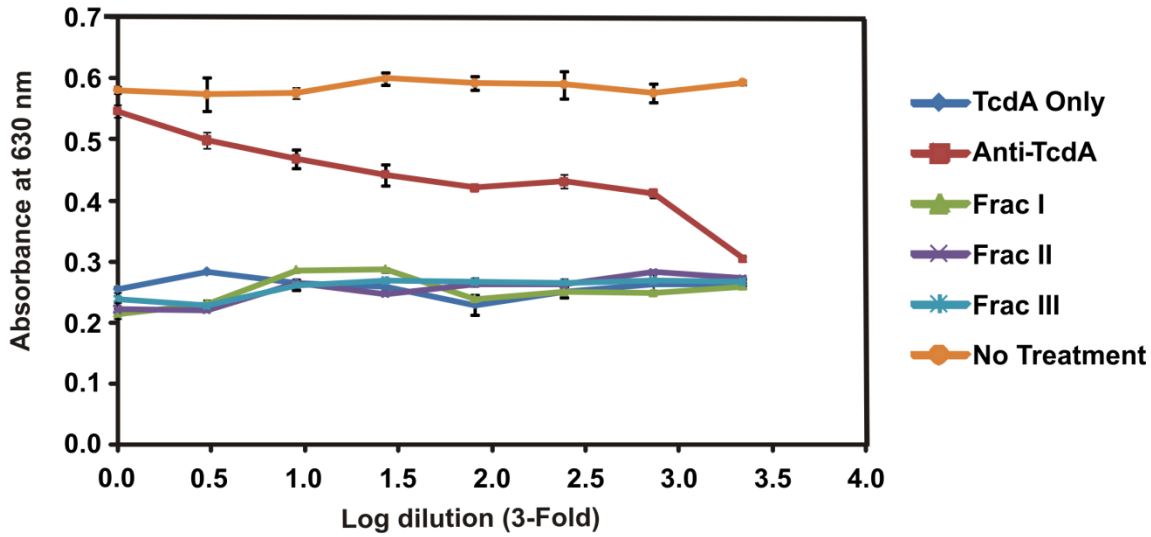
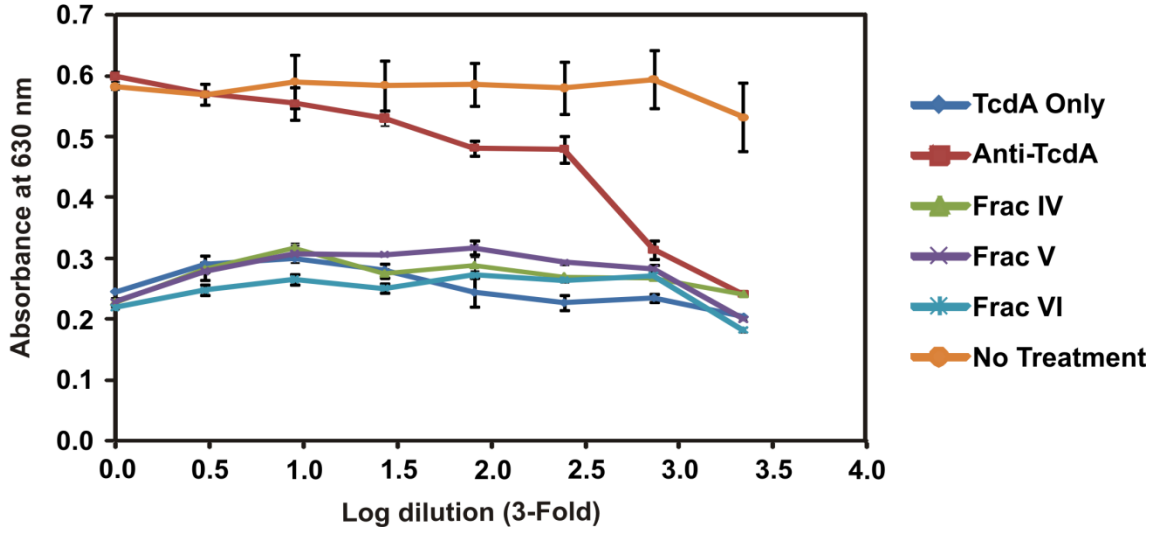


Figure 6

a


Absorbance at 630 nm
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Log dilution (3-Fold)
TcdA Only
Anti-TcdA
Frac I
Frac II
Frac III
No Treatment

b


Absorbance at 630 nm
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Log dilution (3-Fold)
TcdA Only
Anti-TcdA
Frac IV
Frac V
Frac VI
No Treatment

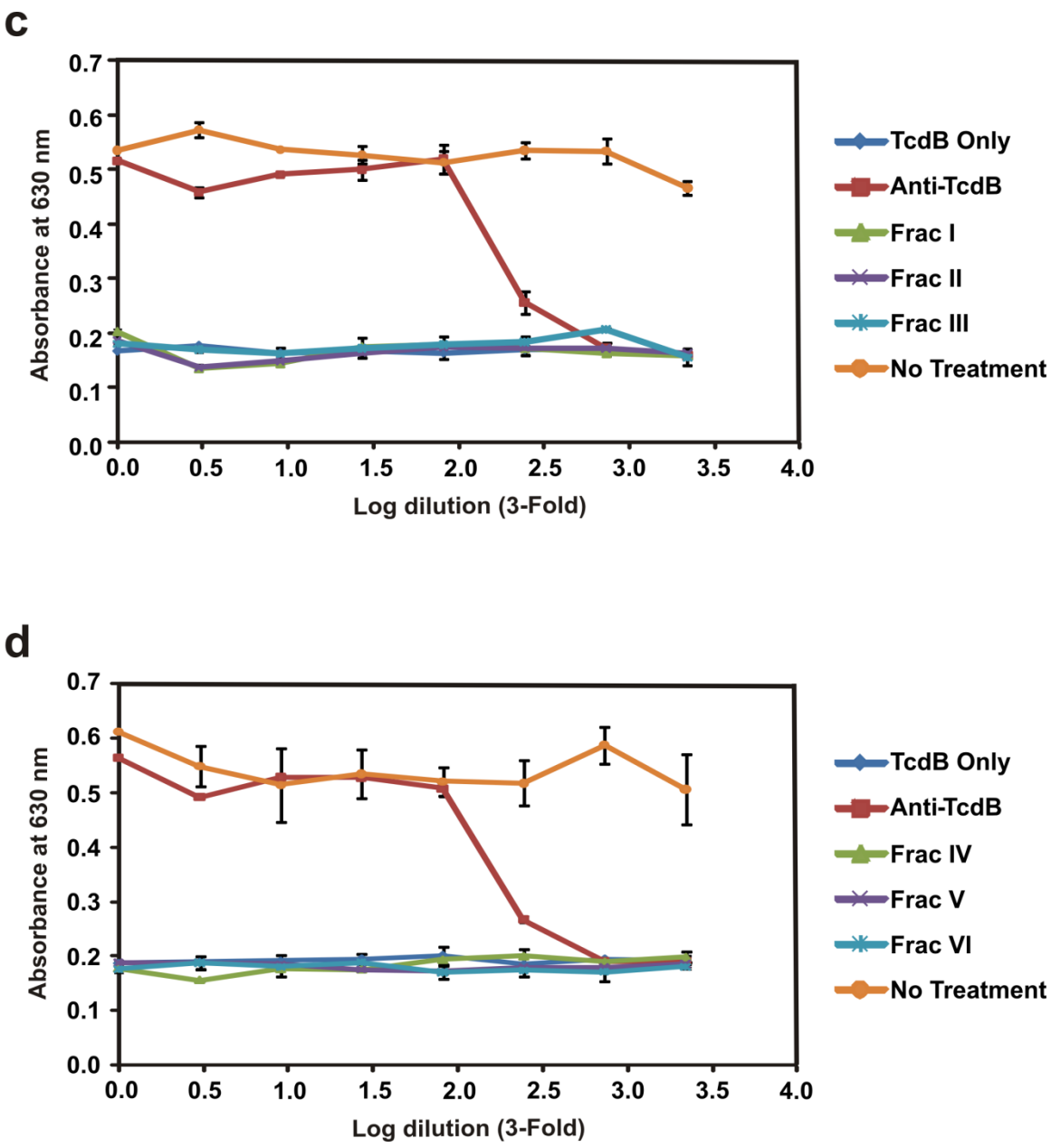

c
Absorbance at 630 nm
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Log dilution (3-Fold)
TcdB Only
Anti-TcdB
Frac I
Frac II
Frac III
No Treatment
d
Absorbance at 630 nm
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Log dilution (3-Fold)
TcdB Only
Anti-TcdB
Frac IV
Frac V
Frac VI
No Treatment

Figure 7

Supplementary Materials

Binding of *Clostridium difficile* Toxins to Human Milk Oligosaccharides

Amr El-Haweit, Elena N. Kitova, Pavel Kitov, Luiz Eugenio, Kenneth K.S. Ng, George L. Mulvey, Tanis C. Dingle, Adam Szpacenko, Glen D. Armstrong and John S. Klassen

Table S1: Mass and composition of the oligosaccharides detected by ES-MS from the HMO fractions.

HMO Fraction	m/z	Mass (measured)	Mass (theoretical)	Monomer composition
Fraction I	643.85	1289.75	1289.42	3Hex:1HexNAc:2NeuAc
	716.90	1435.84	1437.53	5Hex:3HexNAc
	826.48	1654.96	1654.43	4Hex:2HexNAc:2Fuc:1NeuAc, 4Hex:2HexNAc:2NeuAc, 4Hex:2HexNAc:4Fuc
	899.54	1801.08	1800.83	4Hex:2HexNAc:1Fuc:2NeuAc, 6Hex:4HexNAc
	972.07	1946.18	1946.91	4Hex:2HexNAc:3NeuAc 4Hex:2HexNAc:2Fuc:2NeuAc
	1082.17	2166.34	2166.62	5Hex:3HexNAc:1Fuc:2NeuAc, 5Hex:3HexNAc:3Fuc:1NeuAc, 7Hex:5HexNAc
	1155.22	2313.46	2312.41	5Hex:3HexNAc:3NeuAc, 5Hex:3HexNAc:2Fuc:2NeuAc, 5Hex:3HexNAc:4Fuc:1NeuAc, 7Hex:5HexNAc:1Fuc
	1228.27	2458.54	2458.72	5Hex:3HexNAc:1Fuc:3NeuAc, 5Hex:3HexNAc:3Fuc:2NeuAc, 5Hex:3HexNAc:5Fuc:1NeuAc, 7Hex:5HexNAc:1NeuAc 7Hex:5HexNAc:2Fuc
	1288.79	1289.79	1289.42	3Hex:1HexNAc:2NeuAc
Fraction II	643.85	1289.83	1289.42	3Hex:1HexNAc:2NeuAc
	716.90	1435.8	1437.53	5Hex:3HexNAc
	753.93	1509.88	1511.63	4Hex:2HexNAc:1Fuc:1NeuAc**,** 4Hex:2HexNAc:3Fuc
	826.48	1654.96	1654.43	4Hex:2HexNAc:2Fuc:1NeuAc, 4Hex:2HexNAc:2NeuAc, 4Hex:2HexNAc:4Fuc
	997.60	998.61	998.34	3Hex:1HexNAc:1NeuAc, 3Hex:1HexNAc:2Fuc
	1143.70	1144.70	1146.42	3Hex:1HexNAc:1Fuc:1NeuAc
	1508.98	1509.98	1511.63	4Hex:2HexNAc:1Fuc:1NeuAc**,**

				4Hex:2HexNAc:3Fuc
Fraction III	325.16	325.16	326.11	1Hex:1Fuc
	487.26	488.26	488.17	2Hex:1Fuc
	632.35	633.35	633.21	2Hex:2Fuc, 2Hex: 1NeuAc
	643.85	1289.74	1289.42	3Hex:1HexNAc:2NeuAc
	835.50	836.50	838.31	2Hex:1HexNAc:1NeuAc
	997.61	998.61	998.34	3Hex:1HexNAc:1NeuAc, 3Hex:1HexNAc:2Fuc
	1143.72	1144.72	1146.42	3Hex:1HexNAc:1Fuc:1NeuAc
	1217.78	1218.78	1220.54	4Hex:2HexNAc:1Fuc
	1363.89	1364.89	1366.53	4Hex:2HexNAc:2Fuc, 4Hex:2HexNAc:1NeuAc
	1510.00	1511.00	1511.63	4Hex:2HexNAc:1Fuc:1NeuAc**,** 4Hex:2HexNAc:3Fuc
Fraction IV	382.2	383.20	382.12	1Hex:1HexNAc
	632.35	633.35	633.21	2Hex:2Fuc, 2Hex:1NeuAc
	706.41	707.41	707.25	3Hex:1HexNAc
	804.42	782.42	780.32	2Hex:1Fuc:1NeuAc
	852.52	853.52	853.33	3Hex:1HexNAc:1Fuc
	998.62	999.63	998.34	3Hex:1HexNAc:1NeuAc, 3Hex:1HexNAc:2Fuc
	1096.64	1074.64	1074.41	4Hex:2HexNAc
Fraction V	325.16	326.16	326.11	1Hex:1Fuc
	367.18	345.18	344.12	2Hex
	487.26	488.26	488.17	2Hex:1Fuc
	633.36	634.36	633.21	2Hex:2Fuc, 2Hex: 1NeuAc
	997.60	998.60	998.34	3Hex:1HexNAc:1NeuAc, 3Hex:1HexNAc:2Fuc
Fraction VI	341.15	342.15	342.12	2Hex
	377.15	341.15	342.13	2Hex
	683.38	684.38	684.22	2Hex dimer

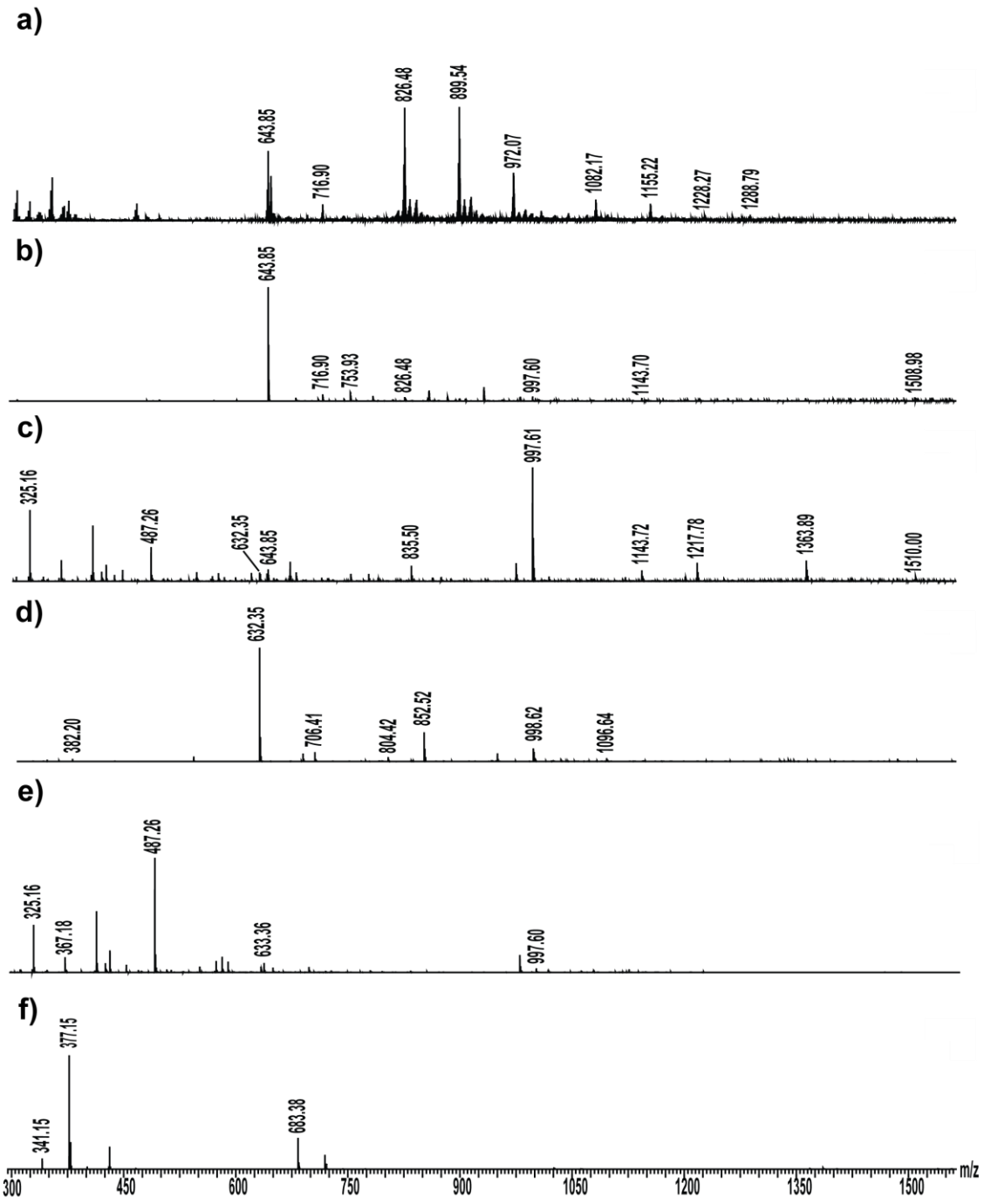


Figure S1: ES mass spectra obtained in negative ion mode with a Synapt HDMS G2 Q-TOF (Waters, Manchester, UK) for (a) HMO fraction I, (b) fraction II, (c) fraction III, (d) fraction IV, (e) fraction V, (f) fraction VI obtained from human milk using size exclusion chromatography separation.

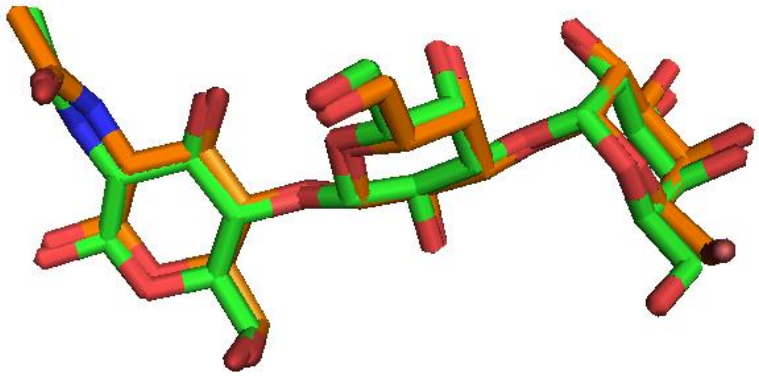


Figure S2. PyMol representation of the minimal energy pose (shown as orange sticks) of Gal(α1-3)Gal(β1-4)GlcNAc(β) obtained following docking with AutoDock Vina and the original pose (shown as green sticks)taken from the crystal structure (Greco et al. 2006).