nature structural & molecular biology

Carbohydrate recognition by *Clostridium difficile* toxin A

Antonio Greco[1,2], Jason G S Ho[1,2], Shuang-Jun Lin[1,3,5], Monica M Palcic[1,3,6], Maja Rupnik[4] & Kenneth K-S Ng[1,2]

Clostridium difficile **TcdA is a large toxin that binds carbohydrates on intestinal epithelial cells. A 2-Å resolution cocrystal structure reveals two molecules of α-Gal-(1,3)-β-Gal-(1,4)-β-GlcNAcO$(CH_2)_8CO_2CH_3$ binding in an extended conformation to TcdA. Residues forming key contacts with the trisaccharides are conserved in all seven putative binding sites in TcdA, suggesting a mode of multivalent binding that may be exploited for the rational design of novel therapeutics.**

Clostridium difficile is a gram-positive anaerobe responsible for hundreds of thousands of nosocomial infections worldwide. Since 2001, a highly virulent clone with a much higher degree of disease incidence, severity and mortality has caused outbreaks in North America and been isolated in Europe (http://www.escmid.org/esgcd)[1]. Although metronidazole or vancomycin is often effective in treating *C. difficile*–associated disease, recurrent infections are common and present a severe therapeutic challenge. Several new approaches to treating *C. difficile*–associated diseases are under development, but a poor understanding of the detailed pathological mechanisms of *C. difficile* limits the effectiveness of all current treatments.

C. difficile produces two major virulence factors, TcdA and TcdB[2,3], which are large (250–308 kDa), single-subunit toxins with multidomain structures. The C-terminal repetitive domain (CRD) binds carbohydrates on colonic epithelial cells as an initial step in pathogenesis. Various oligosaccharides, including the linear B type 2 trisaccharide α-Gal-(1,3)-β-Gal-(1,4)-β-GlcNAc, bind specifically to TcdA, but the native human ligand has not been definitively identified[4,5]. TcdA enters the cell through receptor-mediated endocytosis and disrupts normal signaling pathways necessary for maintaining the cell's cytoskeleton, ultimately leading to inflammation and diarrhea.

Our previously determined structure of a 127-residue C-terminal fragment of TcdA (TcdA-f1) has revealed that the CRD is composed of 31 short repeats (SRs) and 7 long repeats (LRs), with each repeat consisting of a β-hairpin followed by a loop[6]. Uninterrupted stretches of SRs give rise to a left-handed three-fold screw axis with 120° rotations relating β-hairpins in adjacent SRs. In contrast, LRs introduce kinks in which the β-hairpin of the LR is related by a 90° rotation to the β-hairpin of the previous SR[6]. Overall, the domain adopts a β-solenoid fold similar to other bacterial proteins, but the kinks introduced by LRs are unique to TcdA.

Here, we report the crystal structure of the C-terminal 255 residues of TcdA from *C. difficile* strain 48489, toxinotype VI (TcdA-f2) bound to a synthetic derivative of a natural carbohydrate receptor, α-Gal-(1,3)-β-Gal-(1,4)-β-GlcNAcO$(CH_2)_8CO_2CH_3$ (CD-grease). TcdA-f2 corresponds to residues 2456–2710 in the type strain VPI 10463, toxinotype 0, and the sequence of $TcdA_{48489}$-f2 is 94% identical to that in the type strain. Molecular replacement using a search model constructed from TcdA-f1 was used to solve the structure to 2.0-Å resolution (**Supplementary Methods** and **Supplementary Table 1** online). TcdA-f2 contains nine SRs that are punctuated by two LRs (**Fig. 1**).

TcdA-f2 was crystallized at physiological pH and ionic strength in the presence of 5 mM CD-grease. Two binding sites are clearly defined on each of the two molecules of TcdA-f2 in the asymmetric unit (**Figs. 1** and **2**). The entire trisaccharide at each binding site is well-defined by electron density (**Supplementary Fig. 1** online), revealing a single, major conformation similar to that expected in solution[7,8]. For the α-Gal-(1,3)-Gal linkage, $63° < \varphi < 71°$ and $-143° < \psi < 178°$; for the β-Gal-(1,4)-GlcNAc linkage, $-64° < \varphi < -69°$ and $119° < \psi < 123°$ (IUPAC angle definitions). The aliphatic linker attached to the reducing end of the trisaccharide appears disordered.

The carbohydrate-binding sites in TcdA-f2 are shallow troughs consisting of a LR and the hairpin turn of the following SR. The high level of sequence conservation in the LRs and key residues of the following SR suggests that the carbohydrate-binding mode seen in TcdA-f2 is conserved in all seven binding sites identified in TcdA. The interactions between trisaccharides and protein are very similar in both copies of binding site 1 (BS1A and BS1B) and BS2 of chain A (BS2A), but differ in BS2 of chain B (BS2B). Because BS2B may be affected by nearby crystal contacts, only interactions seen in BS1A,

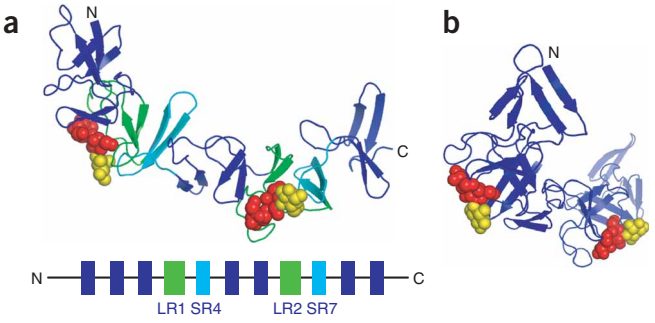

Figure 1 Structure of TcdA-f2. (**a**) Ribbon structure and schematic with SRs (dark blue), SR3 and SR6 (cyan), and LRs (green) highlighted. (**b**) Ribbon representation of TcdA-f2 and bound trisaccharides, viewed from the N terminus. The reducing-end β-GlcNAc is colored yellow.

[1]Alberta Ingenuity Centre for Carbohydrate Science. [2]Department of Biological Sciences, University of Calgary, Calgary, Canada T2N 1N4. [3]Department of Chemistry, University of Alberta, Edmonton, Canada. [4]Institute of Public Health, Maribor, Slovenia. [5]Current address: School of Pharmacy, University of Wisconsin, Madison, Wisconsin, USA. [6]The Carlsberg Laboratory, Valby, Denmark. Correspondence should be addressed to K.K.-S.N. (ngk@ucalgary.ca).

Received 21 December 2005; accepted 9 March 2006; published online 16 April 2006; doi:10.1038/nsmb1084

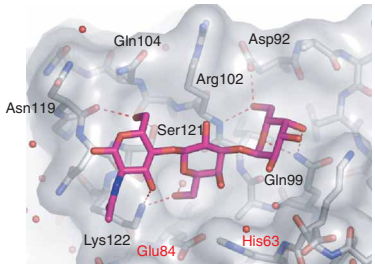

Figure 2 TcdA-f2 bound to CD-grease at BS1A. Hydrogen bonds are drawn as dashed lines.

BS1B and BS2A are discussed below. For simplicity, BS1A is described; equivalent residues in BS2 would be numbered $n + 91$.

About 350 Å^2 or 50% of the carbohydrate surface is buried upon binding. The 6-OH of β-galactose accepts a hydrogen from Lys122 and donates a hydrogen to a water molecule at the bottom of the binding pocket (**Fig. 2**), which is also seen in the unliganded structure[6]. The B face of β-galactose packs against the apolar proximal portions of Arg102 and Ser121. The α-galactose at the nonreducing end lies in a pocket formed by highly conserved residues in the loop after the β-hairpin of the LR. Its hydrophobic B face packs against Ile101, and the 6-OH forms hydrogen bonds with Asp92 and Arg102, whereas the 4-OH accepts a hydrogen from Gln99. The O5 ring oxygen also accepts hydrogen bonds from Gln99 and a water molecule. The binding pocket is fairly open and may accommodate variations at the nonreducing end of the trisaccharide, such as α-GalNAc-(1,3), β-GalNAc-(1,3) or β-GlcNAc-(1,3) residues, which have previously been shown to bind TcdA[5].

The reducing-end β-GlcNAc lies in a shallow surface pocket formed primarily by the β-turn portion of the SR after the LR. The 3-OH accepts a hydrogen from Lys122, which also donates a hydrogen to the β-galactose 6-OH. The β-GlcNAc 6-OH also accepts a hydrogen from a main chain carbonyl oxygen atom, and the B-face packs against residues 120 and 121. Notably, the reducing end of the trisaccharide extends away from the protein, which allows space for attached lipid or protein parts of the native receptor. The TcdA-f2–carbohydrate complex also suggests a possible binding mode for Le^x and Le^y[4]. If the β-Gal-(1,4)-β-GlcNAc moiety binds TcdA as seen for CD-grease, there is sufficient space to accommodate the α-fucose residues attached to GlcNAc or galactose in Le^x and Le^y.

Several previous studies support the functional importance of carbohydrate binding in TcdA-f2. First, diethylpyrocarbonate modification of histidine residues in TcdA specifically abolishes cytotoxicity and receptor-binding activities[9]. The TcdA-f2 complex shows that His63 is situated within 5 Å of the central galactose residue (**Fig. 2**). This residue forms a three-way hydrogen bond network with Glu84 and Tyr56 at the conserved interface between each LR and the preceding SR[6]. Six of the nine histidine residues in the TcdA CRD lie near carbohydrate-binding sites, and the modification of histidine is expected to disrupt six of the seven binding sites. Second, antibodies raised against TIDGKKYYFN inhibit receptor binding and cytotoxicity in TcdA[10]. This sequence is similar to that found in the SR preceding each putative carbohydrate-binding site in TcdA. Antibodies bound to this sequence would probably block carbohydrate binding. Finally, polymeric resins conjugated to α-Gal-(1,2)-β-Gal-(1,4)-β-Glc inhibit receptor binding and cytotoxicity in TcdA[11,12].

The locations of carbohydrate-binding sites in TcdA-f2 suggest how TcdA may form multivalent interactions with intestinal cell-surface carbohydrates. Structural modeling indicates that the TcdA CRD adopts an elongated, serpentine shape in which multiple binding sites may be presented on the same face of the structure (**Supplementary Fig. 2** online). The reducing ends of bound trisaccharides point away from the protein, minimizing steric interference with glycolipid and glycoprotein ligands. Adjacent binding sites are expected to lie on the same face if they are separated by three SRs. However, adjacent binding sites separated by four or five SRs are presented on surfaces that are 120° apart. As a result, multiple binding sites separated by 30-300 Å may interact with carbohydrates simultaneously, but not all binding sites can bind at the same time. Multivalent binding spaced over large distances has been seen in many other lectin-cell surface interactions[13].

Several novel therapeutic approaches interfering with the activities of toxins from *C. difficile* address limitations in current treatments with antibiotics. Passive immunization with antisera targeting the TcdA and TcdB CRDs reduces damage to the intestine caused by *C. difficile* infections[14]. Polyanionic or carbohydrate-containing resins also bind toxins and neutralize their pathogenic effects[11,12,15,16]. Structural information on TcdA-carbohydrate interactions provides a rational basis for improving the effectiveness of these and other promising new therapeutic approaches.

Accession codes. Protein Data Bank: Coordinates and structure factors have been deposited with accession code 2G7C.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank I. Barrette-Ng for helpful discussions. This work was supported by the Alberta Ingenuity Centre for Carbohydrate Science (M.M.P. and K.K.-S.N.), the Canadian Institutes of Health Research (K.K.-S.N.) and the Alberta Heritage Foundation for Medical Research (AHFMR; to K.K.-S.N.). J.G.S.H. was supported by a fellowship from the Alberta Ingenuity Fund. M.R. was supported by the Slovenian research Agency (project J1-6456). Diffraction data were collected at beamline 8.3.1 of the Advanced Light Source (ALS) at Lawrence Berkeley Lab under an agreement with the Alberta Synchrotron Institute (ASI). The ALS is operated by the US Department of Energy and supported by the US National Institutes of Health. Beamline 8.3.1 was funded by the US National Science Foundation, the University of California and Henry Wheeler. The ASI synchrotron access program is supported by the Alberta Science and Research Authority and the AHFMR.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/nsmb/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Bartlett, J.G. & Perl, T.M. *N. Engl. J. Med.* **353**, 2503–2505 (2005).
2. Rupnik, M. *et al. J. Med. Microbiol.* **54**, 113–117 (2005).
3. Voth, D.E. & Ballard, J.D. *Clin. Microbiol. Rev.* **18**, 247–263 (2005).
4. Tucker, K.D. & Wilkins, T.D. *Infect. Immun.* **59**, 73–78 (1991).
5. Teneberg, S. *et al. Glycobiology* **6**, 599–609 (1996).
6. Ho, J.G., Greco, A., Rupnik, M. & Ng, K.K. *Proc. Natl. Acad. Sci. USA* **102**, 18373–18378 (2005).
7. Corzana, F. *et al. Glycobiology* **12**, 241–250 (2002).
8. Yuriev, E., Farrugia, W., Scott, A.M. & Ramsland, P.A. *Immunol. Cell Biol.* **83**, 709–717 (2005).
9. Roberts, A.K. & Shone, C.C. *Toxicon* **39**, 325–333 (2001).
10. Wren, B.W., Russell, R.R. & Tabaqchali, S. *Infect. Immun.* **59**, 3151–3155 (1991).
11. Heerze, L.D., Kelm, M.A., Talbot, J.A. & Armstrong, G.D. *J. Infect. Dis.* **169**, 1291–1296 (1994).
12. Castagliuolo, I., LaMont, J.T., Qiu, B., Nikulasson, S.T. & Pothoulakis, C. *Gastroenterology* **111**, 433–438 (1996).
13. Weis, W.I. & Drickamer, K. *Annu. Rev. Biochem.* **65**, 441–473 (1996).
14. Kelly, C.P. *et al. Antimicrob. Agents Chemother.* **40**, 373–379 (1996).
15. Kurtz, C.B. *et al. Antimicrob. Agents Chemother.* **45**, 2340–2347 (2001).
16. Braunlin, W. *et al. Biophys. J.* **87**, 534–539 (2004).