Structural basis of carbohydrate recognition in
Clostridium difficile Toxin A

Antonio Greco1,2, Jason G.S. Ho1,2, Shuang-Jun Lin1,3,5, Monica M. Palec1,3,6,
Maja Rupnik4, and Kenneth K.-S. Ng1,2,‡

1Alberta Ingenuity Centre for Carbohydrate Science
2University of Calgary, Department of Biological Sciences, Calgary, Alberta, Canada
3University of Alberta, Department of Chemistry, Edmonton, Alberta, Canada
4Institute of Public Health Maribor, Maribor, Slovenia
5Current address: School of Pharmacy, University of Wisconsin, Madison, Wisconsin, 53705
6The Carlsberg Laboratory, Valby, Denmark

Supplementary Methods

Cloning, Expression, and Purification
To generate an expression clone for TcdA-f2, PCR (forward primer = 5’ G GAA TTC CAT ATG CAC CAT CAC CAT CAC TCA AAA GCA GTT ACC GGA ATG CG and reverse primer = 5’ CGG GAT CCC CTA TC C TTA GCC ATA TAT CCC AGG GGC TTT TAC TCC) was used to amplify the coding region for residues 2456-2710 of TcdA (C. difficile strain 48489; toxinotype VI; numbering according to strain VPI 10463; toxinotype 0) using clone pA3-484891. Following restriction enzyme digestion with Ndel and BamHI, this fragment was ligated into pET-3a (Novagen) and transformed into E. coli JM109. Dideoxy chain-termination sequencing was used to verify the sequence of the expression clone.

TcdA-f01 and TcdA-f2 were expressed in E. coli BL21 (DE3) pLysS following induction with 0.5 mM IPTG and growth in LB medium at 25°C for 18 hr. Cells from 1 L of culture were harvested by centrifugation and resuspended in 35 mL lysis buffer (100 mM sodium phosphate, 200 mM sodium chloride, 5 mM imidazole, pH 8.0). The cells were treated with 0.5 mg DNase I and 0.1 mM PMSF for 10 minutes, lysed by sonication, and clarified by centrifugation. The clarified extract was chromatographed on Nickel-NTA-Sepharose (1x5 cm column, Qiagen), yielding 15 mL at 3.0 mg/mL. Following dialysis against 20 mM MES, pH 6.0, 100 mM NaCl, 5% glycerol, 0.5 mM EDTA,
TcdA-f2 was concentrated to 3 mg/ml using Vivaspin 15R concentrators, 5000 MWCO (Vivascience).

**Synthesis of carbohydrates**

αGal(1,3)βGal(1,4)GlcNAcO(CH₂)₈CO₂CH₃ was enzymatically synthesized using recombinant α(1-3)galactosyltransferase (8 units) with 8-methoxycarbonyloctyl N-acetyllactosamine acceptor (68 mg) and UDP-galactose dipotassium salt donor (112 mg) as previously described for purified calf thymus enzyme². αGal(1,3)βGal(1,4)βGlcO(CH₂)₈CO₂CH₃ was enzymatically synthesized using recombinant α(1-3)galactosyltransferase with 8-methoxycarbonyloctyl lactoside acceptor and UDP-galactose donor as previously described³.

**Crystallization and data collection**

A lyophilized sample of αGal(1,3)βGal(1,4)βGlcNAcO(CH₂)₈CO₂CH₃ was dissolved at 100 mM in water and mixed with TcdA-f2 to yield a protein:carbohydrate mixture consisting of 2.7 mg/mL protein and 10 mM carbohydrate. Crystals were grown by the vapour diffusion method (1 L protein:carbohydrate mixture + 2 L reservoir equilibrated against 0.5 mL reservoir) at 21°C. Initial crystallization conditions were obtained from sparse matrix screens (Index-HT and Crystal-HT, Hampton Research). The optimized reservoir solution used for the crystallization of TcdA-f2 was 6% PEG 3350, 100 mM BisTris-Cl pH 7.0, 5% glycerol. Diffraction data were measured from a needle-like crystal (~0.2 X 0.03 X 0.03 mm) that was briefly transferred into a modified reservoir solution containing 20% glycerol and 10 mM αGal(1,3)βGal(1,4)βGlcNAcO(CH₂)₈CO₂CH₃ before being flash-cooled in a nitrogen gas stream at ~100 K. Diffraction data were initially measured using a MAR 345 image plate and X-rays produced with a rotating copper anode (Rigaku RUH3R). Higher resolution data were measured using an ADSC Quantum-315 CCD detector at the Advanced Light Source on beamline 8.3.1. Data were processed and scaled using DENZO, Scalepack and programs from CCP4 (version 5.0.2)⁵,⁶. The space group was determined to be P2₁ by autoindexing, data reduction and examining systematic absences. Crystallographic statistics are summarized in Table 1.
The solvent content of the crystals was calculated to be 56% ($V_m = 2.80$) if two copies of TcdA-f2 were present in the asymmetric unit. The structure of TcdA-f1 (PDB code 2F6E) was used as the search model for molecular replacement calculations carried out using PHASER\textsuperscript{7}. The cross-rotation and translation functions both gave two prominent solutions (LL = 77.8/75.8 and 108.9/104.6). Sequence analysis revealed a pattern of SRs and LRs that allowed the construction of a reasonable model for the unknown portion of the structure. Rigid-body refinement ($R = 0.40$, resolution = 20-3.2 Å) and positional/temperature factor refinement were carried out using CNS\textsuperscript{8} and REFMAC\textsuperscript{9}. Xfit\textsuperscript{10} was used for model building. Difference electron density maps clearly identified the location and conformation of two trisaccharide molecules bound to each copy of TcdA-f2 in the asymmetric unit. Subsequently, a data set extending to higher resolution was measured using the same crystal (T = 100K, $\lambda = 1.115\AA$, ALS beamline 8.3.1; Supplementary Table 1). And additional rounds of refinement and model building were performed. Residues 13-257 in chain A and residues 8-261 in chain B are well-ordered and have been modeled. The model for chain B is only missing the N-terminal Met, the His\textsubscript{6} tag and the C-terminal Gly residue. 89.8 % of residues lie in the most favoured regions of the Ramachandran plot and no residues are in disallowed regions (regions defined by PROCHECK\textsuperscript{11}). Additional checks on model geometry were performed using WHATCHECK\textsuperscript{12}. Data quality and refinement statistics are given in Table 1. Figures were prepared with PyMOL\textsuperscript{13}. 

3
REFERENCES