Structure-Function Study of Large Clostridial Toxins and their Glucosyltransferase Domains

By

Joseph W. Alvin

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

In partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Chemical and Physical Biology

August 11, 2017

Nashville, Tennessee

Approved:

D. Borden Lacy, Ph.D.

Albert Beth, Ph.D.

Nicholas J. Reiter, Ph.D.

Timothy L. Cover, M.D.

Acknowledgements

This work was made possible by support from Vanderbilt University including the Biomedical Research Education and Training office, Molecular Biophysics Training Program, Center for Structural Biology, National Institutes of Health, and Department of Veterans Affairs. My journey in research has always been supported by my family and friends, with inspiration from my grandfather and the mentors who taught me. I would not have succeeded without the love, care, and understanding of my own family of mammals, especially Tara.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

APD	Autoprocessing Domain
BGTD	Glucosyltransferase domain of TcdB
CROPs	Combined repetitive oligopeptides
CV	Column Volumes
GAP	GTPase activating protein
GAL	Galactose
GDI	Guanine nucleotide dissociation inhibitor
GDP	Guanine diphosphate
GEF	Guanine nucleotide exchange factor
GST	Glutathione S-transferase
GTD	Glycosyltransferase domain
GTP	Guanine triphosphate
GTPase	Guanine triphosphotase
kpsi	Thousand pounds per square inch
LCT	Large clostridial toxin
NDP	Nucleotide diphosphate
pAzF	*para*-Arylazidephenylalanine
pBzF	*para*-Benzophenonephenylalanine
PDB	Protein database
PVDF	Polyvinylidene difluoride
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography

TcdA	*C. difficile* toxin A
TcdB	*C. difficile* toxin B
Tcnα	*C. novyi* toxin α
TcsH	*C. sordellii* toxin H
TcsL	*C. sordellii* toxin L
TpeL	*C. perfringens* toxin L
UPG	Uridine diphosphate glucose
U2F	Uridine 2-deoxy-2-fluoroglucose
UDP	Uridine diphosphate
UDP-GlcNAc	Uridine diphosphate N-acetyl glucosamine

Chapter 1

Structure and Function of Large Clostridial Toxins

1-1 Introduction

The clostridium genus is comprised of a highly diverse set of anaerobic, gram-positive, spore-forming organisms. Nevertheless, some share the capacity to produce a homologous family of proteins called the large clostridial toxins (LCTs) (Table 1-1). In each instance, the production of toxins is thought to contribute to the link each organism has with disease.

Clostridium difficile is a bacterial pathogen that causes nearly half a million infections in the Unites States each year[1,2]. It infects the human colon and can cause diarrhea, pseudomembranous colitis and, in some cases, death. As the major virulence factors in *C. difficile* infection (CDI), TcdA and TcdB have become the prototypical members of the LCT family. The toxins are responsible for the diarrhea, inflammation, and colonic tissue damage that occur with CDI. While strains containing only TcdB have been associated with severe disease in both human and experimental animal infections, most clinical isolates express both toxins. The respective role that each toxin plays in the context of infection is an active area of inquiry[3].

C. sordellii infections are most often associated with gynecological events, intravenous drug use, or musculoskeletal allografts[4]. While rare, these infections can cause a profound toxic shock syndrome characterized by treatment-refractory hypotension, extreme leukocytosis, and visceral effusions. Infections are associated with a high mortality rate. The major toxins for this organism are two members of the

Table 1-1. Comparison of amino acid sequence identity

	TcdA	TcdB	TcsH	TcsL	Tcnα	TpeL
TcdA						
TcdB	48%					
TcsH	78%	48%				
TcsL	48%	76%	49%			
Tcnα	31%	30%	31%	30%		
TpeL	41%	41%	42%	42%	33%	

LCT family, TcsH and TcsL. Notably, TcsH shares greater sequence identity with TcdA of *C. difficile* than with TcsL, and a similar relationship exists between TcsL and TcdB (Table 1). The toxins have been shown to be important in animal models and in the evaluation of human infections, although there are a number of clinical isolates that only express TcsL.

Infections by *C. novyi* are typically the result of soil contact with open wounds or injection drug use with soil-contamination[4]. Symptoms of disease can include wound-associated gas gangrene or a clostridial toxic shock syndrome that is similar to what is observed in *C. sordellii* infections. Vaccination with Tcnα toxoid in guinea pigs significantly improved survival after challenge with *C. novyi* spores[5], and strains cured

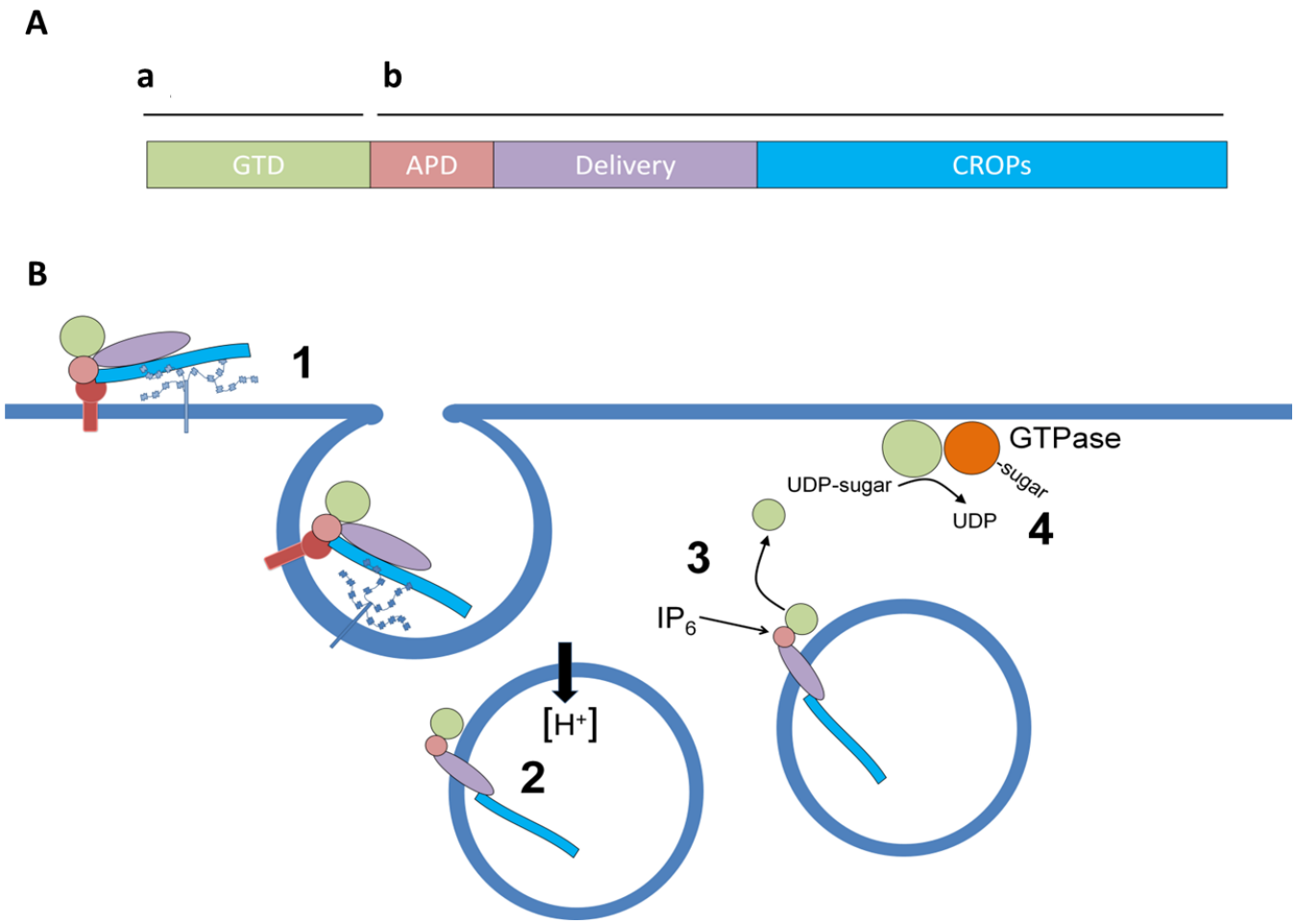

Figure 1-1. Overview of domain organization and model of LCT intoxication. LCTs are comprised of (N to C-terminus) a glycosyltransferase domain (GTD), autoprocessing domain (APD), delivery or pore-forming domain (PFD), and combined repetitive oligopeptides (CROPs). Step 1; the CROPs is thought to help mediate LCT cell binding through interactions with cell surface glycoproteins and glycolipids. The LCT can also interact with a host receptor that contributes to specificity and endocytosis. Step 2; as the endosome acidifies, the delivery domain inserts into the lipid bilayer, forming a pore. The GTD (and likely the APD) pass through this pore into the host cytosol. Step 3; IP6 binds to the APD, initiating autoproteolysis and the release of the GTD from the endosome. Step 4; the GTD localizes at membranes and glycosylates specific subsets of host GTPases. This glycosylation disrupts downstream signaling of the GTPase pathways and results in both cytopathic and apoptotic effects[3].

of the Tcnα-encoding bacteriophage show reduced virulence in animal models[4]. These observations underscore the importance of Tcnα in disease pathogenesis.

Although *C. perfringens* is an etiological agent in gas gangrene and foodborne illnesses, the role (if any) for TpeL in human disease is unclear. The presence of TpeL has been linked to an increase in the frequency and size of enteric necrotic lesions in broiler chicks, however[6]. This suggests that TpeL may play an important role in necrotic enteritis of chickens and other agriculturally significant birds. Still, much of the current understanding of TpeL structure and function has come from a comparative analysis with other members or the LCT family. *C. difficile* produces two large, homologous proteins which are responsible for cytotoxicity in humans and other animals. The toxins from *C. difficile* (TcdA and TcdB) are homologous to other cytotoxins produced in related *Clostridia spp.* including *C. sordellii* (TcsL and TcsH), *C. novyi* (Tcnα), and *C. perfringens* (TpeL). Homology between these LCTs is evident given the previously discussed cross-reactivity of *C. sordellii* antiserum to *C. difficile* cytotoxins. The LCTs are large single-chain proteins ranging from 191 to 308 kDa in molecular weight. They are considered AB-toxins in that a 'B domain' is responsible for the binding and delivery of an enzymatic 'A domain' into the cytosol of the host cell (Figure 1-1A). In each LCT, the 'A domain' is a glycosyltransferase (GTD) that targets members of the Rho and Ras GTPase families. The 'B domain' can be divided into three regions: the autoprocessing domain (APD), delivery domain (DD), and combined repetitive oligopeptide repeats (CROPs) domains. These discrete structural domains contribute to a multi-step mechanism of intoxication that includes 1) receptor binding and endocytosis, 2)

translocation and delivery of the GTD across the endosomal membrane 3) autoprocessing and 4) glycosyltransfer (Figure 1-1B).

1-2 Receptor binding and the CROPs

With the exception of TpeL, the carboxy-terminal region of the LCTs encodes a series of repetitive sequences that together form the CROPs domain[7]. The repeats are composed of multiple 19-24 amino acid short repeats interspersed with 31 amino acid long repeats (Figure 1-2A). In electron microscopy (EM) images of the TcdA holotoxin, the CROPs is evident as an extended structure (Figure 1-2B). Crystal structures of fragments from the CROPs domains of TcdA suggest an extended β-solenoid structure where the short and long repeats come together to form a vertex (Figure 1-2C). In TcdA, this vertex has been shown to bind an α-Gal-(1,3)-β-Gal-(1,4)-β-GlcNAc trisaccharide (Figure 1-2D).

Historically, the CROPs has been thought to represent the receptor-binding domain of the LCTs[7]. Much of the evidence in support of this view has come from the study of TcdA and the capacity of the TcdA CROPs to bind carbohydrates. The sugar binding properties of the TcdB CROPs appear to be distinct and have not been investigated in detail for the other LCTs. TcdA can bind to the human I, X, and Y blood antigens as well as a human glycosphingolipid, and all have a core β-Gal-(1,4)-β-GlcNAc structure. In addition to the capacity to bind carboyhydrates, excess TcdA CROPs domain competes with TcdA holotoxin for cell binding. Lastly, the CROPs domains are highly antigenic, and antibodies against the TcdA and TcdB CROPs can neutralize the effects of their respective toxins *in vitro.* In one example, such antibodies have shown efficacy in protecting people against recurrent CDI[8]. The mechanism of

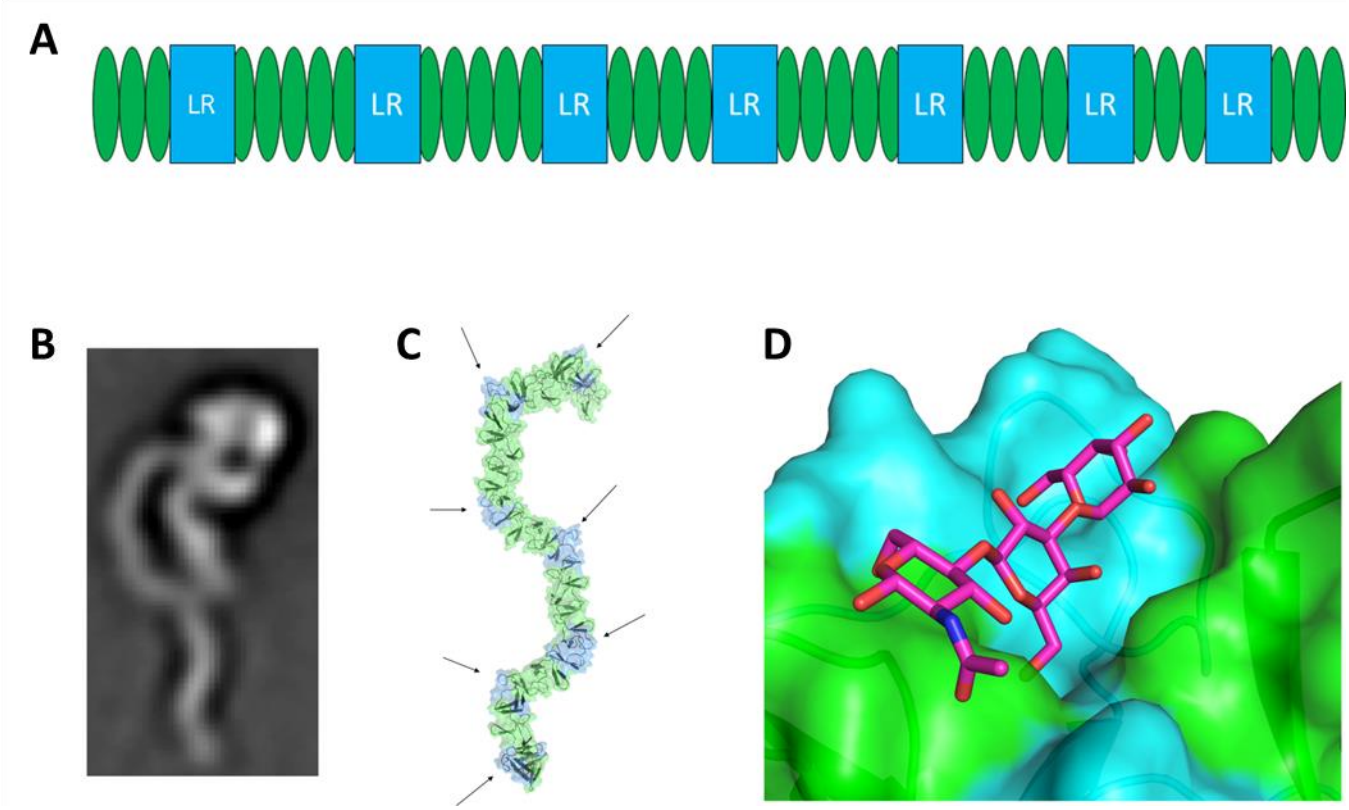

Figure 1-2. Structural organization of the CROPs domain. A) TcdA-CROPs consists of seven long repeat (LR) sections, separated by varying numbers of short repeats (SR). **B)** Electron microscopy shows TcdA-CROPs adopts a sinusoidal structure, similar to what was observed in **C)** a model derived from the crystal structure of a TcdA CROPs fragment. In TcdA, each of the LR vertices can bind carbohydrates (indicated by arrows). **D)** a crystal structure of a CROPs SR-LR-SR fragment bound to a α-Gal-(1,3)-β-Gal-(1,4)- β-GlcNAc trisaccharide[3].

this protective effect is not currently known, but could result from a block in receptor binding.

More recent studies have indicated that LCT receptor-binding function is not limited to the CROPs domain. Truncations of TcdA and TcdB that lack the CROPs domains are still capable of intoxicating cells, and TpeL lacks a CROPs domain entirely[7]. LDL receptor-related protein 1 (LRP1) was shown to act as a TpeL receptor[9], confirming the idea that specific binding and entry of the toxin does not rely on the presence of a CROPs domain. Similarly, chondroitin sulfate proteoglycan 4 (CSPG4) and poliovirus receptor-like protein 3 (PVRL3) were recently reported as TcdB receptors, and both bind in regions outside of the CROPs[10,11].

In light of these observations, the role of the CROPs domain in LCT function is not entirely clear. It could be that it contributes to but is not essential for binding, consistent with a proposed two-receptor binding model [9]. It is also possible that the importance of the CROPs to receptor binding varies among the LCTs. For example, in contrast to what was observed with TcdA, the application of recombinant TcdB CROPs does not compete for TcdB binding. Finally, it could be that the CROPs is involved in processes outside of the host entry mechanism, for example, interaction with the producing bacterium or other components of the infected host. These possibilities merit investigation in the context of the lifestyle and pathogenesis associated with the different LCT-producing bacteria.

1-3 Delivery into the cytosol

LCT receptor-binding is followed by receptor-mediated endocytosis (Figure 1-1B). As the endosome matures, the pH becomes more acidic. The low pH triggers a structural change in the LCT allowing it to insert into the endosomal membrane and form a pore. The pore can then serve as a conduit for the delivery of the glucosyltransferase domain (and perhaps also the autoprocessing domain) into the cytosol. The requirement for an acidic pH is supported by the observations that bafilomycin, an inhibitor of the vacuolar H^+-ATPase, protects against the cytopathic and/or cytotoxic effects of TcdA, TcdB, and TcsL. Pore formation can also be triggered at the cell surface by acidification of the extracellular medium and quantified by a measure of cellular Rb^+ release. Other indications of pore formation and pH-dependent conformational changes have come from EM, fluorescence, planar lipid bilayer and liposome fluorescence release assays[7].

The LCT central region that separates the APD from the CROPs has been termed the pore-forming or delivery domain based on the presence of conserved hydrophobic sequences that are thought to insert into the membrane with low pH. The recent crystal structure of a TcdA construct containing the first 1832 amino acids of TcdA ($TcdA_{1832}$) provides the first atomic model of the delivery domain in its soluble form (Figure 1-3A)[12]. The domain begins as a three-helix bundle at the GTD-APD interface (residues 767-841, Figure 1-3A, blue) which leads into a small globular sub-domain (residues 850-1025, Figure 1-3A, red). This sub-domain is then followed by an extended 'hydrophobic helical stretch' containing four α-helices (1026-1135, Figure 1-3A, yellow) which wrap around a series of β-sheet structures (1136-1802, Figure 1-3A, purple). The β-sheet

structures end at the base of the APD and are thus positioned to transition into the junction with the CROPs domain. An EM analysis of TcdA holotoxin structure at neutral and low pH indicates that the junction between the delivery domain and the CROPs is a point of flexibility (Figure 1-3B)[7,13].

While the molecular details of how pore formation and translocation occur are currently unclear, mutational studies suggest that residues within the globular sub-domain and the hydrophobic helical stretch are important for pore formation[14] . Many of the residues within this hydrophobic stretch are conserved across the LCT family and have been demonstrated to be important for the formation of membrane pores in the context of TcdA and TcdB. The presence of a surface turn at the base of the helical stretch which is important for both pore formation and cytotoxicity and which is conserved across the entire LCT family (Figure 1-3A) suggests an attractive target for broad spectrum neutralization by antibodies or other therapeutic molecules[12].

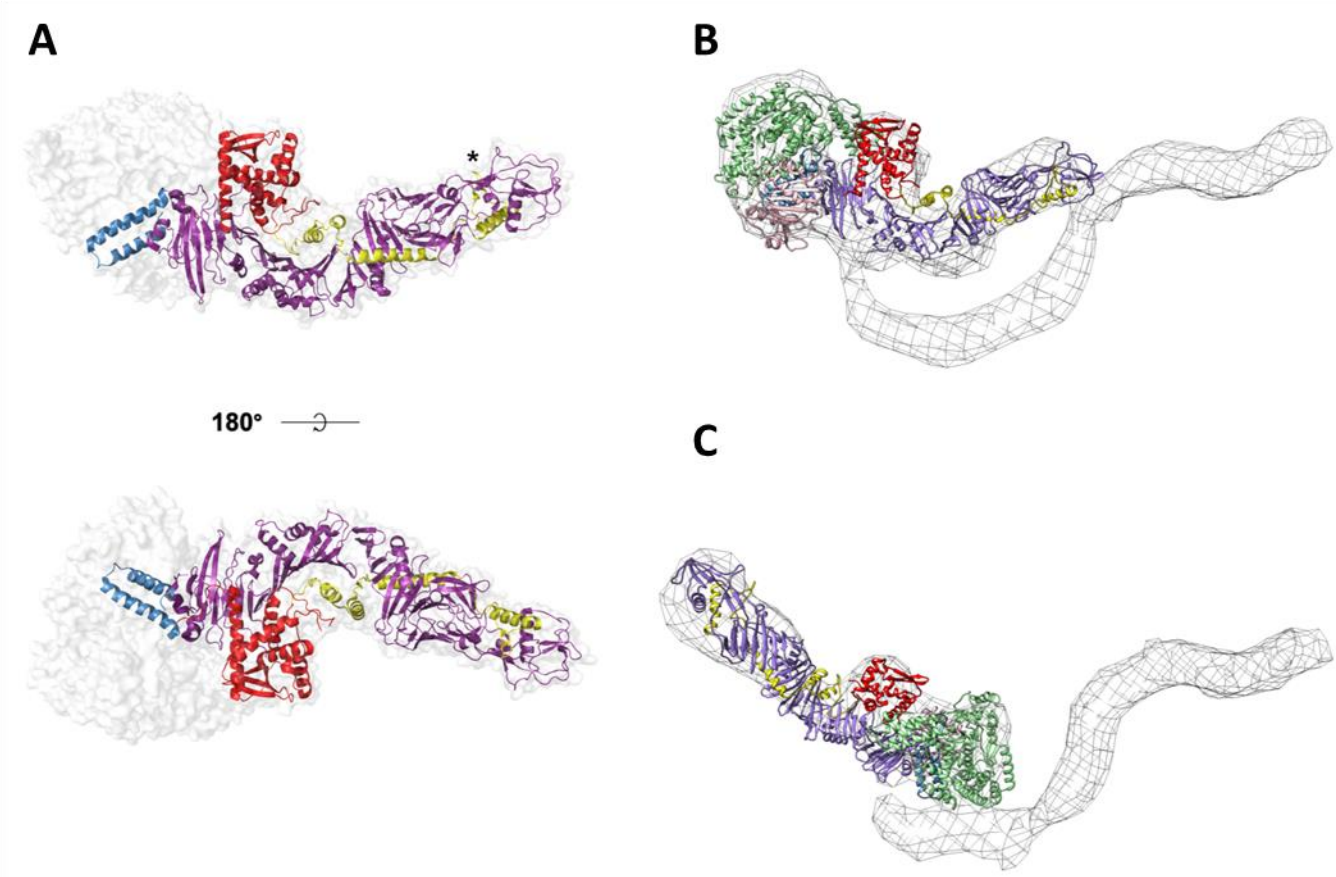

Figure 1-3. Electron densities of TcdA crystal structure and negative stain EM. A) The crystal structure of TcdA 1-1832 is shown as a transparent surface model with the delivery domain highlighted in color. Key features of the delivery domain include a three-helix bundle (blue), a globular sub-domain (red), and the α-helical hydrophobic stretch (yellow). The conserved surface loop is indicated with an asterisk. The rest of the domain is colored in purple and is primarily composed of β-sheet structures. The TcdA 1-1832 structure with the GTD in green, the APD in pink, and the delivery domain colored as in (A) was docked within TcdA EM structures obtained at **B)** neutral and **C)** acidic conditions[3].

1-4 Autoprocessing

Following the delivery of the GTD across the endosomal membrane, the autoprocessing domain (APD) mediates cleavage at a conserved leucine residue to release the GTD into the cytosol[7]. *In vitro*, this cleavage is initiated by the binding of inositol hexakisphosphate (IP_6), a small molecule found at high concentrations within the eukaryotic cell cytosol. Intracellular concentrations of IP_6 in mammalian cells have been reported in ranges of 10 μM to 1 mM. Isothermal titration calorimetry experiments have been used to calculate the affinity of IP_6 for multiple LCT APDs in the form of an equilibrium dissociation constant (K_D): 4.4-5.1 μM for TcdB, 2.1 to 2.8 μM for TcsL, and 6.7 to 8.9 μM for Tcnα[15]. These data, along with *in vitro* experiments demonstrating holotoxin autoprocessing at IP_6 concentrations as low as 0.1 μM, support the idea that IP6 acts as the small molecule activator in physiological contexts.

IP6 binds the APD and causes an allosteric change that is propagated through a central β-flap region which separates the electropositive IP6 binding site from the active site of the APD[7]. Structures of the TcdA and TcdB APDs bound to IP6 have revealed highly similar structures where the IP6 is coordinated by several positively charged residues. While efforts to crystallize the isolated ADP of either toxin in the absence of IP6 have been unsuccessful, a comparison of the apo-APD structure (crystallized in the context of $TcdA_{1832}$) with that of IP6-bound APD reveals the significant rearrangements that occur with IP6 binding[12]. Most notably, the domain's C-terminal helix (TcdA 766-778) uncoils at either end, and the central β-flap (TcdA 746-765) rotates towards the IP6 binding site and away from the enzyme active site by ~90° (Figure 1-4A). One effect of this rearrangement is that the side chains of four positively charged residues move into

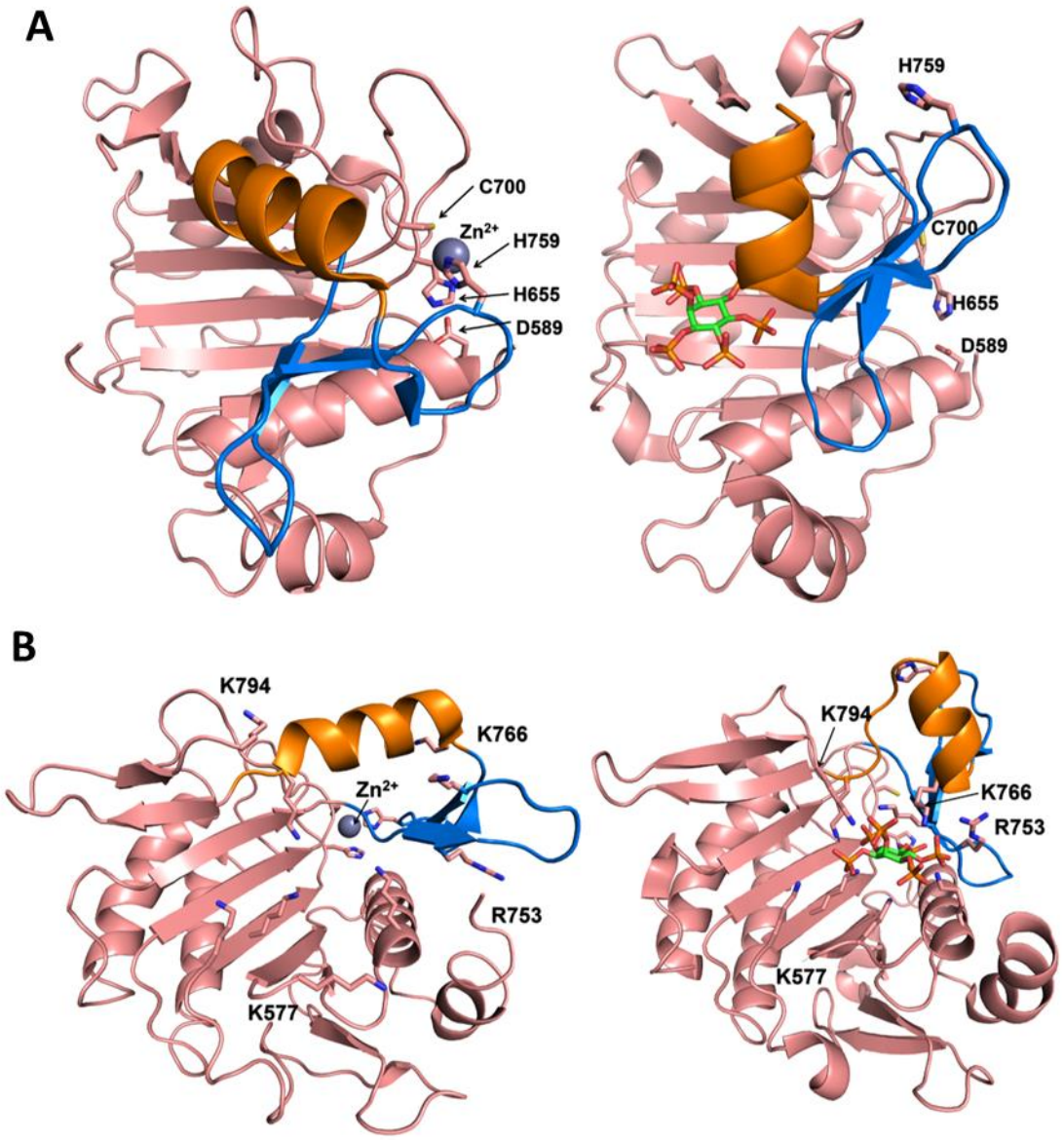

Figure 1-4. TcdA-APD conformational changes induced by IP6 binding A) comparison of the TcdA APD in the absence of IP6 (left, the *apo* state) versus the IP6-bound structure (right) reveals conformational changes in the β-flap (blue) and a C-terminal α-helix. H759 coordinates a Zn^{2+} ion in the active site of the apo-APD but moves out of the active site upon IP6 binding. Residues C700, H655, and D589 are located within the active site in both structures and have been shown to have a role in autoprocessing activity. **B)** an alternate view of the apo (left) and IP6-bound structures (right) highlights the positively charged residues (K577, K766, R753, and K794) that undergo conformational changes in order to bind IP6[3].

the IP6 binding site to bind the IP6 phosphate groups (Figure 1-4B). In addition, the H759 residue, located at the tip of the β-flap, moves from being a part of the enzyme active site to a location 19 Å away (Figure 1-4A). Mutation of H759 in TcdA to alanine (or H757 in TcdB) results in a protein whose efficiency of autoprocessing no longer depends on IP6 concentration[12]. These data are consistent with a model where the structural changes imposed by IP6 binding trigger allosteric changes in the enzyme active site that permit autoprocessing.

The active site has been described as a catalytic triad in that it has a conserved trio of cysteine, histidine, and aspartic acid residues[7]. Mutation of any one of these residues (D589, H655, or C700 in TcdA) results in a loss or significant defect in autoprocessing and has led the field to describe the APD as a cysteine protease. While crystal structures of the TcdA and TcdB APDs in the presence of IP6 did not indicate the expected arrangement of these three residues (Figure 1-4A, right), the absence of substrate made it hard to evaluate the proposed mechanism of activity.

The $TcdA_{1832}$ crystal structure led to the unexpected discovery that both TcdA and TcdB have a zinc bound to the cysteine of their APD active sites (Figure 1-4), and this zinc is required for autoprocessing activity[12]. In light of these observations it is unclear whether autoprocessing results from an activated cysteine, zinc-activated water, or some other nucleophile. Structures in the presence of substrate could help resolve this question. In the meantime, the more general reference to an autoprocessing domain (rather than a cysteine protease domain) seems warranted. The presumed function of the APD is to release the GTD from the endosomal surface, thereby allowing it to migrate to GTPase substrates located at the plasma membrane (Figure 1-1B).

Studies in TcdA and TcdB have indicated that while autoprocessing is important for the kinetics of cytopathic responses, it is not essential[16–18]. TcdA and TcdB toxin mutants that are deficient in autoprocessing will still cause phenotypic changes suggesting that some GTPase substrates are associated with the endosomal membrane and/or that toxin containing endosomes are recycled to the cell surface. It is interesting to note that when a TcsL autoprocessing mutant is applied to cells, the rates of Rac1 and Ras modification differ[19]. This is consistent with known differences in Rac and Ras localization and suggests that autoprocessing may be more important in LCTs that target Ras GTPase proteins that are exclusively localized at the plasma membrane.

1-5 Glycosyltransfer

The LCTs encode an N-terminal 63 kDa GTD that can transfer a sugar onto small Rho and Ras family GTPases[7]. The specificity for different GTPase family members can differ among the LCTs[20] (Table 1-2). In general, the glycosyltransferase

Table 1-2. Selectivity of LCT GTDs for disparate GTPases. This table represents the ability (denoted by '+') of LCT GTDs to glycosylate various GTPases. Asterisks represent cases where LCTs are capable of modifying a different set of substrates or where the GTPase is not a preferred substrate.

	Rho(ABC)	RhoG	Rac1	Cdc42	(HKN) Ras	R-Ras	Ral(ABC)	Rap1	Rap2
TcdA	+	+	+	+	*		*		*
TcdB	+	+	+	+			*	*	*
TcsH	+	+	+	+	*	*	*		*
TcsL	*	*	+	*	+	+	+	+	+
Tcnα	+	*	+	+					
TpeL			+		+	*	*	+	

mechanism involves the nucleophilic attack of an activated nucleotide sugar which results in the transfer of the sugar on to the starting nucleophile. In the LCT's, the nucleophile is the conserved threonine in the Switch I region of the GTPase. The modification of the threonine prevents the exchange of GDP for GTP and the capacity of the GTPase to bind regulatory molecules and downstream effectors and therefore, disrupts multiple downstream signaling pathways. The modification of the Rac1 GTPase has been linked to changes in the actin cytoskeletal structure and a cytopathic rounding effect. The capacity of TcsL to inactivate Ras has also been linked to TcsL-induced cell death. Other phenotypic effects of toxin-induced GTPase inactivation include the disruption of tight junctions and the production of cytokines.

The LCTs use either UDP-glucose or UDP-GlcNac as the glycosyl donor. UDP-glucose serves as the source of sugar for TcdA, TcdB, TcsH, and TcsL, Tcnα uses UDP-GlcNAc, and TpeL can use either UDP-glucose or UDP-GlcNAc as the co-substrate[21]. The TpeL capacity to bind either UDP-glucose or UDP-GlcNAc is based on the primary sequence at amino acid positions 383-385. Where TcdA, TcdB and TcsL have an INQ sequence, TpeL and Tcnα have ANQ or SNA, respectively. Substrate discrimination depends on steric clashes between the sugar C2 moieties and side chains of this trio. Mutations in Tcnα from 385-SNA to 385-INQ changes the donor substrate from UDP-GlcNAc to UDP-glucose. Similarly, an A383I mutation in TpeL results in a preference for UDP-glucose[22]. Measurements of intracellular UDP-glucose demonstrate concentrations approaching 100 μM[23]. This exceeds the K_m for TcdB-GTD by >10-fold and implies the rate-limiting factor for GTD substrate engagement is the local concentration of the target GTPases[24].

Crystal structures are available for the TcdA, TcdB, TcsL, and Tcnα GTDs [7,22]. All are similar in that they share a Rossmann fold at their core and four helical sub-domains that are thought to contribute to target specificity (Figure 1-5A). The core is similar to what has been observed for other members of the glycosyltransferase type A family of enzymes. A conserved DXD within this core (D285 and D287 in TcdA) is important for the coordination of Mn^{2+}, and there are two conserved tryptophans (W101 and W519 in TcdA) that bind UDP and the glycosidic oxygen, respectively

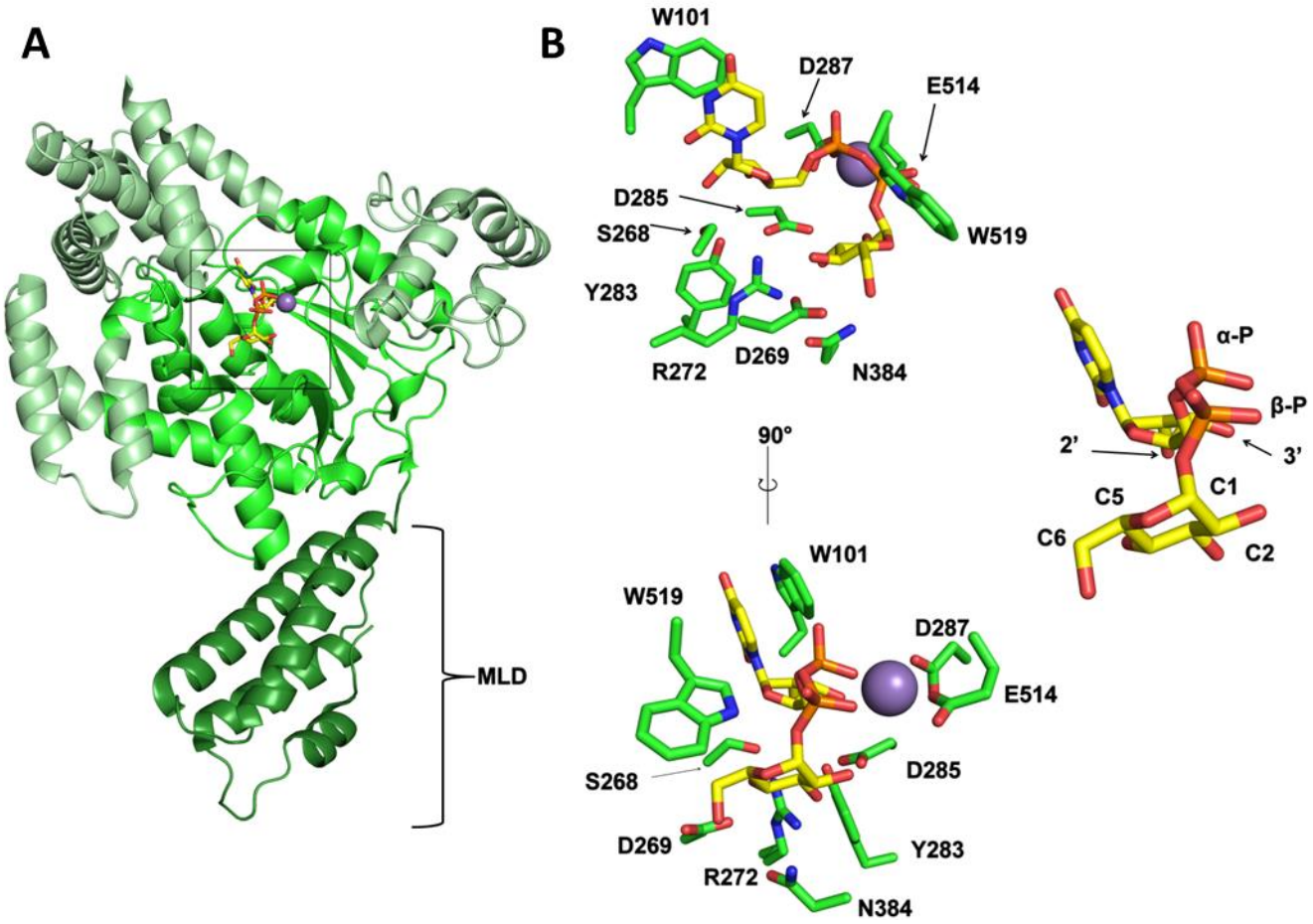

Figure 1-5. The TcdA glucosyltransferase domain (GTD) bound to UDP-glucose and Mn^{2+}. A) The LCT GTD has a Rossmann fold at its core (bright green) that is involved in the coordination of UDP-glucose and catalysis. Four α-helical sub-domains emerge from this core: three in light green that are thought to contribute to GTPase binding and the N-terminal membrane localization domain (MLD) (dark green) that allows the GTD to bind the eukaryotic cell membrane. **B)** UDP-glucose structure (right) and functionally important residues within the GTD active site (left). W102 coordinates the uridine base of UDP-glucose through π-π stacking. Y283 and D285 form hydrogen bonds with the 2' and 3' hydroxyl groups on the ribose. R272 and D285 help bind glucose by interacting with the C3 hydroxyl group. The acidic residues D287 and E514 chelate the manganese cofactor along with the diphosphate backbone and two water molecules (not pictured). Both W519 and N384 play important roles in catalyzing the proposed S_Ni rearrangement. N384 is thought to stabilize the transition state where partial bonds exist between the glucose C1 hydroxyl, β-phosphate, and GTPase threonine Cβ hydroxyl groups. The indole nitrogen of W519 also plays a role in catalysis through interactions with the oxygen linking the β-phosphate and glucose C1[3].

(Figure 1-5B). Mutation of these conserved residues can lead to toxins that are highly attenuated in their glycosyltransferase activities.

The mechanism of how the sugar is transferred is not fully understood. While some glycosyltransferases use an inverting mechanism, one where the nucleophilic attack by the accepting atom results in a change of stereochemistry at the sugar's anomeric bond, the LCTs do not appear to function in this way. Work from Aktories and colleagues has shown that the stereochemistry of the anomeric bond is retained in the course of the reaction[24]. Multiple mechanisms that could explain the retention in stereochemistry have been put forward, but none have been fully accepted. For example, a double displacement reaction could explain the retention in stereochemistry but implies the existence of a covalent sugar-GTD intermediate, for which there is no evidence. Another retaining mechanism could involve S_Ni (nucleophilic substitution with internal return). An S_Ni mechanism permits simultaneous bond lysis and formation through a short-lived oxocarbenium transition state rather than a covalent intermediate. TcdB has been proposed to undergo an S_Ni mechanism where a conserved asparagine (N384) stabilizes a ring-like transition state and bonds are simultaneously broken and formed between UDP-glucose and a sterically constrained threonine. This hypothesis is supported by the observation that TcdB N384A has a 1200-fold reduction in glycosyltransferase activity[25]. A crystal structure of an LCT GTD in complex with GTPase would be useful in resolving these mechanistic questions.

The four additional sub-domains that emerge from the Rossmann fold core are unique to the LCTs (Figure 1-5A). While it is assumed that one or more of these sub-domains contribute to the LCT specificity for Rho and Ras family GTPases, the

molecular basis of this specificity and how GTPase specificity differs among LCT family members is not currently understood. Studies have demonstrated the effects of several mutations in TcdB-GTD which diminish its capacity to modify some GTPases[24].

A series of charged residues are located between helix 16 and helix 17, which comprise residues ~440-480. Exchanging the sequence of helix 17 from TcsL into TcdB significantly impaired glucosylation of Rho, Rac, and Cdc42 by TcdB-GTD. However, this exchange was not sufficient to enable TcdB-GTD to glucosylate the TcsL substrates Ras or RalA. When TcsL was modified to include the entire helix 17 from TcdB, modification of Ras and Ral was not observed. However, a small amount of glucosylated RhoA was found, suggesting that some elements of TcdB-GTD helix 17 contribute to the capacity to recognize RhoA[25].

Most GTPases are modified at a C-terminal cysteine with an extended hydrophobic moiety such as myristoyl, palmitoyl, or prenyl groups. In the absence of regulatory proteins, these nonpolar groups localize GTPases to lipid membranes, where some function as regulators of the cytoskeleton and some function in receptor signaling. One of the four sub-domain structures in the GTD is a four helix bundle that is formed by the first ~90 residues of each LCT[7]. This sub-domain acts as a membrane localization domain (MLD) with specificity for lipids found on the inner leaflet of the plasma membrane domain. The MLD is thought to orient the GTD in a position to modify the membrane-associated GTPases. Mutation of conserved residues extending from the tip of the MLD four helix bundle have been shown to impair TcsL glycosylation of substrates[19,26]. While the TcsL GTD preferentially binds membranes enriched in phosphatidylserine, the TcdB GTD has equivalent rates of glucosylation in liposomes

containing phosphatidylserine or phosphatidylglycerol. Thus, the MLDs also play a role in determining each LCT's specificity for GTPase substrates.

1-6 LCT function

The functional outcomes of LCT intoxication can vary depending on what cell type and tissue is exposed to toxin. As an enteric pathogen, the LCTs from *C. difficile* (TcdA and TcdB) are thought to first encounter the epithelial cells of the colonic epithelium. The same may be true for TpeL in the context of necrotic enteritis. In contrast, the *C. sordellii* and *C. novyi* toxins are thought to encounter endothelial cells and cause a vascular leak that contributes to edema, hypotension, and hemoconcentration[6]. Endothelial and epithelial cells function as the basis for fluid, gas, nutrient, and waste exchange in nearly every tissue and are thus, essential in maintaining homeostasis.

When LCTs intoxicate endothelial or epithelial cells, they cause changes in cytoskeletal structure that result in a cytopathic effect (CPE). The concomitant disruption of tight junctions permits fluid secretion and edema, as well as the release of immunomodulatory cytokines and chemokines that call innate immune cells to the site of infection. The toxins can also promote cell death by either apoptotic or necrotic mechanisms depending on the toxin and the toxin concentration.

1-7 Glycosylation-dependent cytopathic and cytotoxic effects

The cytopathic effect of cell rounding is a hallmark of LCT intoxication and is the result of GTPase glycosylation. The modification of the Rac1 GTPase has been linked to changes in the actin cytoskeletal structure and results in an arborized *difficile*--type (D-type) CPE observed in response to TcdA and many forms of TcdB[27]. TcsL can modify Ras family GTPases to cause a spindle-like *sordellii*-type (S-type) cytopathic effect. These cells are characterized by intact focal adhesions *despite* a loss of central actin organization. Interestingly some $TcdA^-/B^+$ strains of *C. difficile* cause S-type CPEs, similar to the effects of TcsL (*C. difficile* strains 8864, J9965, WA151, 1470, and ES130)[28]. The GTD's of these TcdB producing strains are capable of modifying broader arrays of GTPases, including those of the Ras family. It is currently unclear whether these discrete cytopathies play distinct roles in the context of infection.

A key consequence of the cytopathic effect is a cellular arrest at the G_2/M checkpoint. Intoxication by TcdA and TcsH leads to a rapid increase in RhoB expression in multiple cell types [24]. RhoB is a pro-apoptotic GTPase which functions through interactions with Protein Kinase B (Akt) and the p38 MAPK pathways. TcsL and TpeL are thought to cause apoptosis through inhibition of GTPase-mediated (e.g., Ras) regulation of the Akt/PI3K survival pathway[29,30]. In sum, all LCT's have the capacity to stimulate apoptotic pathways of cell death as a result of their glycosyltransferase activities.

1-8 Glycosylation-independent necrosis

In addition to the apoptotic events that can be observed after 24-72 hours of intoxication, TcdB can stimulate a rapid necrotic form of cell death when applied to cells or colonic tissue explants at concentrations of 100 pM and higher[12,31]. Notably, the necrotic mechanism does not depend on either the autoprocessing or glucosyltransferase activities of the toxin. Instead, TcdB can stimulate the assembly of the NADPH oxidase (NOX) complex on the surface of epithelial cell endosomes which leads to the production of reactive oxygen species (ROS) that go on to kill the cell[32]. This mechanism of necrotic cell death may explain the presence of necrotic lesions in the colons of patients with severe forms of CDI and the link between TcdB and disease in animal models of infection[33]. While it is known that TcdA does not stimulate the NOX response, there are not yet reports of whether this effect has been observed in response to other LCT family members.

1-9 LCT mechanism as a tool for novel therapeutic discovery

While antibiotics are a first step in almost all strategies to treat LCT-associated bacterial infections, the LCTs themselves are attractive targets for the development of novel therapeutic and prevention strategies. Many studies have shown that toxoid vaccines of *C. difficile* or *C. sordellii* LCTs or infusion with purified antibodies provide significant protection in animal infections and some human studies[34,35]. In keeping with the theme of this chapter, we focus our discussion below on therapeutic concepts that are being driven by the understanding of LCT toxin structure and mechanism[36].

The mechanism of action depicted in Figure 1-1 suggests several mechanistic steps that could be targeted to block intoxication: receptor binding and entry, acidification of the endosome, pore formation, autoprocessing, and glycosylation of GTPases. In concept, one could block the initial interaction of the LCT with its cell surface receptor. This is the presumed mechanism of humanized monoclonal antibodies that have been shown to reduce the recurrence of CDI in clinical trials[8]. Antibodies from this study are known to bind the TcdA and TcdB CROPs. However, the current set of TcdB receptors--CSPG4, PVRL3, FZD do not bind the TcdB CROPs. In the FZD study, several TcdB constructs of amino acids 1114–1835 demonstrated binding to the FZD cysteine-rich domains (CRD)[37]. This section contains the canonical separation of the delivery from the CROPs domain. Based on the $TcdA_{1-1832}$ crystal structure and electron microscopy the precise cut-off between the delivery and CROPs domains is unclear. However, when TcdA undergoes the acid-induced conformational change, this domain transition area seems to act as a pivot. Further study is clearly needed to fully understand the interactions between toxins, receptors, CROPs, and cell surfaces.

Further efforts to understand the mechanism of action for neutralizing antibodies and to specifically target LCT-receptor interactions are on-going. Several efforts at chemical screening have revealed compounds that link to known mechanistic steps of intoxication. For example, a cell based phenotypic screen for small molecules that block TcdB toxicity identified inhibitors of endosome maturation including concanamycin A (an endosome acidification inhibitor)[38]. The bile salts methyl cholate and taurocholic acid also protect against TcdB-mediated toxicity[39,40]. While one could speculate that

this results from an impact on membrane structure that affects receptor binding and/or internalization, there is also evidence that methyl cholate binds directly to TcdB to cause a conformational change[39,40].

Efforts to limit the autoprocessing activities of the toxins have centered around the conserved cysteine that serves either as nucleophile or in the coordination of zinc. An electrophilic peptide mimetic was able to bind the cysteine and limit the cytopathic effects of the toxin[41]. Similarly, the small molecule ebselen was found to be a potent inhibitor of TcdB autoprocessing through modification of the APD active site cysteine[42]. Importantly, ebselen also showed efficacy in protecting against CDI disease pathology *in vivo*. While ebselen is known to have many activities in cellular function (including a disruption of NOX complex assembly)[43], the possibility that ebselen could be acting though inhibition of the autoprocessing activity is intriguing. *In vitro* studies indicate that the autoprocessing activity is not strictly required for TcdA and TcdB function [16–18], but the role of autoprocessing in an *in vivo* context has not been evaluated.

Another target for inhibition is the glycosyltransferase activity of the LCT GTDs. Natural products such as castanospermine, swainsoline, and 1-deoxynojirimycin are broad inhibitors of glycosidases and glycosyltransferases[44]. Castanospermine was used to obtain a crystal structure of TcsL and appeared to occupy the equivalent position of glucose or GlcNAc from similar LCT-GTD structures. Injection of castanospermine prior to LCT application prevented cytopathic and cytotoxic effects of TcsL and TcdB. Castanospermine was found to have K_i values of ~100 μM for Rac1 glucosylation and UDP-glucose hydrolysis by TcsL-GTD, but was less effective at inhibiting TcdB-GTD (K_i ~400 μM; 120 μM, respectively)[24]. Interestingly, the K_i against these GTDs was

significantly improved when UDP was included, which suggests occupation of both UDP and hexose sites enhances complex stability. Recently, a common plant flavonoid—phloretin—was shown to protect cells against TcdA and TcdB-induced cell death [39]. Subsequent *in vitro* studies found that phloretin inhibits the glucosyltransferase activity of the TcdA/B-GTDs with an IC_{50} ~ 2 μM[39]. This effect was shown to be noncompetitive with UDP-glucose and had a much smaller effect on UDP-glucose hydrolysis (IC_{50} > 200 μM). These data suggest the flavonoid acts allosterically on the GTD or GTD-GTPase complex, significantly reducing the rate of glucosylation without affecting UDP-glucose binding.

The importance of the GTDs in these toxins from *C. difficile* and related organisms is the basis of my thesis work. In an effort to further our understanding, I sought to study the structural characteristics of the TcdA/B-GTD using a combination of sequence comparison and crosslinking, substrates, inhibitors, and crystallography. The key questions driving this work are: 1) how do GTDs discriminate GTPase substrates; 2) what is the catalytic mechanism for glucosyltransfer; and 3) how does the small molecule apigenin selectively inhibit glucosyltransfer catalysis.

Chapter 2

Towards an understanding of Glycosyltransferase specificity and LCT variation

2-1 Introduction

Since its identification in the 1980s, *C. difficile* infections have gradually become a significant healthcare burden in the US and abroad. The pathology of infections is dependent on the secretion of two large toxins named TcdA and TcdB. Typically, disease-causing strains of *C. difficile* produce both toxins. However, as early as 1991 researchers isolated strains which produced TcdB, but had undetectable DNA encoding TcdA[45]. While the clinical relevance of a bi-toxigenic ($TcdA^+/TcdB^+$) versus monotoxigenic ($TcdA^-/TcdB^+$) *C. difficile* is unclear, studies show that variations in TcdB sequence alter the cytotoxic effects. The cytopathic effect (CPE) of cell rounding is a hallmark of intoxication and is the result of GTPase glycosylation. The modification of the Rac1 GTPase by TcdA and TcdB has been linked to changes in the actin cytoskeletal structure These cytoskeletal disruptions result in a phenotype of arborized cells described as a *difficile*--type (D-type) CPE[27]. The TcsL toxin from *C. sordellii* shares 76% sequence identity with TcdB, causes a distinct CPE where cells fail to maintain distal focal adhesions and causing completely rounded cells. TcsL can modify Ras family GTPases and this is thought to contribute to this cytological differences

observed in the *sordellii*-type (S-type) CPE. Interestingly some $TcdA^-/B^+$ strains of *C. difficile* cause S-type CPEs, similar to the effects of TcsL (*C. difficile* strains 8864, J9965, WA151, 1470, and ES130)[28]. The GTD's of these TcdB producing strains are capable of modifying broader arrays of GTPases, including those of the Ras family. Interestingly, the GTD's of these $TcdB_{var}$ toxins are also able to modify GTPases of the Ras family. Canonically, TcdB-GTD modifies RhoA, Rac1, Cdc42, but not members of the Ras family GTPases. In contrast, TcsL-GTD primarily modifies Rac1, (H, K, N, R)-Ras, Ral and Rap GTPases, but not Rho-family members[46,47]. Studies of TcdB-GTD from strains 1470 and 8864 demonstrate their ability to glucosylate R-Ras *in vitro*[48]. Furthermore, the 1470 and 8864 TcdB-GTDs were deficient in their ability to inactivate RhoA and Cdc42. Together, these data show that some TcdB variants can cause discrete effects on cells. Amino acid differences between canonical and variant TcdB could help in determining which residues are used by TcdB-GTD to bind and modify GTPases.

Our collaborator, Maja Rupnik, shared a collection of 19 *tcdB* sequences obtained from clinical *C. difficile* isolates. These sequences were combined with sequences from canonical *C. diffiicle* strains such as 630 and 10463 and sequences from $TcdA^-/TcdB^+$ strains. These DNA sequences were translated to amino acid sequence and subsequently aligned using the Clustal Kalign algorithm[49]. Protein sequences from *C. difficile* TcdB (10463 and 8864) and *C. sordellii* TcsL were submitted to ESPript 3 to identify variable positions[50]. The surface distribution of these variable sites is shown in Figure 2-1. Variable amino acid positions are colored blue on the crystal structure of $TcdB\text{-}GTD_{10463}$. Interestingly, the bulk of variable sites within the

TcdB-GTD are located on the concave face, which affords access to the UDP-glucose in the active site and facilitates the transfer of glucose from UDP-glucose to target GTPases (Figure 2-1). By integrating the strain-specific GTPase glucosylation experiments and models of GTD structure, I sought to identify significant residues related to GTPase specificity.

2-2 Toxin variation and GTPase specificity

We generated a model of the TcdB-GTD from strain 8864 based on the crystal structure of TcdB-GTD$_{10463}$. We then compared the electrostatic predictions from this model to the crystal structures of TcdB$_{10463}$ (PDB 2BVL) and TcsL$_{6018}$ (PDB 2VK9) (Figure 2-2)[51]. While the specific interactions between LCT-GTDs and GTPases have not been defined, previous work had shown that residues within GTD helix 16 and 17 (residues 440-480) play important roles in determining GTPase specificity[52]. The boxed areas in Figure 2-2 represent the surface of alpha helices 16-17, which are more positively charged in both TcsL$_{6018}$ and TcdB$_{8864}$. This similarity between TcdB$_{8864}$ and TcsL$_{6018}$ is located at residue 449 in alpha helix 16 (Figure 2-2, inset bold text). All strains within groups two and four (sordellii-like CPEs) have lysine at this position, compared to groups one and three which have glutamate. It has been demonstrated that this mutation—E449K—is sufficient to impair glucosylation of RhoA and Cdc42[52]. This suggests a plausible explanation for the divergent CPEs across the *C. difficile* strains in this study. However, it is important to clarify that while the E449K mutation impacts glucosylation of these GTPases, it is not sufficient to determine substrate specificity. Typically, substituting acidic amino acids with basic residues is used to eliminate salt bridges and/or drive similarly charged groups apart. While the E449K

mutation does reverse charge, $TcdB_{10463}$ contains a lysine at positon 452. K452 is within hydrogen-bonding distance to E449, which could result in a polar, but non-charged surface. This neutral surface of $TcdB_{10463}$ may facilitate interaction with RhoA and Cdc42, whereas the positively charged surface of $TcdB_{8864}$ and TcsL does not. We then sought to use chemical crosslinking between the proteins in an attempt to test these hypotheses.

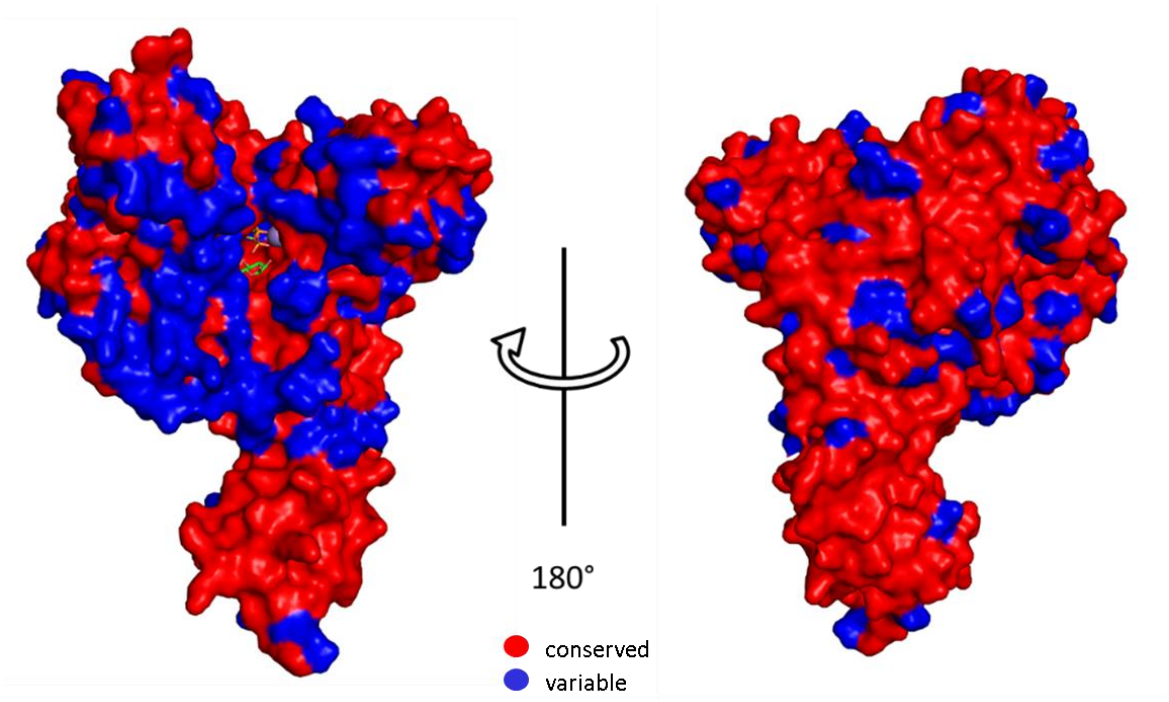

Figure 2-1. Surface topology of variation in TcdB-GTD across *C. difficile* strains. Amino acid sequences from 19 *C. difficile* strains were aligned using the Clustal Kalign algorithm. Positions where amino acids vary is shown are blue while strictly conserved amino acids are shown in red. This figure was generated using PyMol and PDB 2BVM[77].

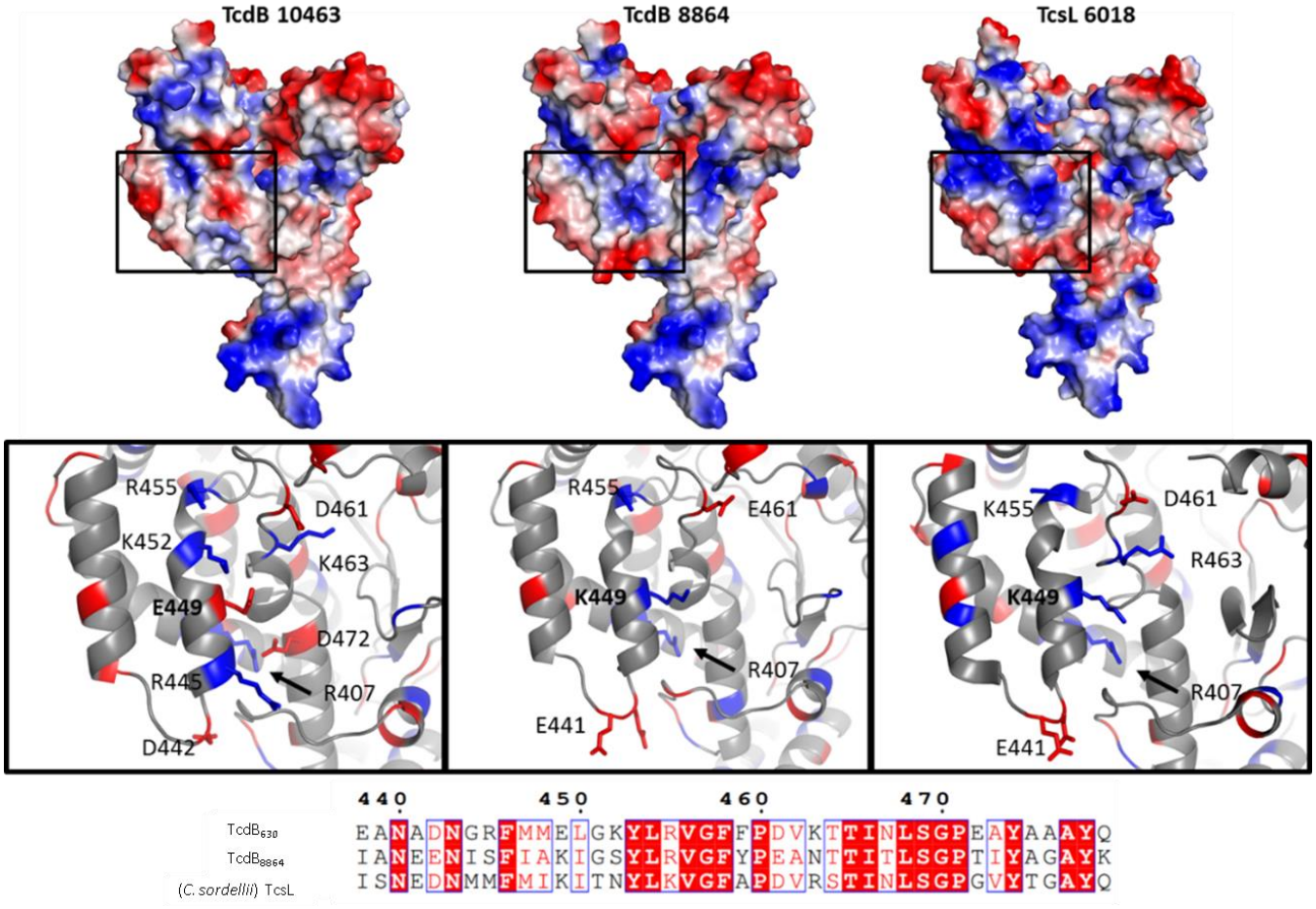

Figure 2-2. Electrostatic plots of *C. difficile* TcdB-GTDs and TcsL-GTD. The amino acids (1-544) from *C. difficile* strain 8864 were submitted to SWISS-MODEL, which produced a model of the GTD based on the structure of TcdB-GTD from *C. difficile* strain 10463 (PDB 2BVM). Sequences from *C. difficile* (10463 and 8864) and *C. sordellii* were aligned with Clustal and submitted to ESPript 3 to create the colored alignment. Strictly conserved residues have a red background, while similar mutations have red text, and nonsimilar mutations have black text. Vacuum electrostatics were generated using PyMol.

2-3 Crosslinking as a method to define the GTD/GTPase interface

Identifying the residues responsible for GTPase specificity within LCT-GTDs would significantly enhance our understanding of these toxins. A co-crystal structure of these proteins would demonstrate all residues involved in the protein-protein interaction. Given the high homology across the LCT-GTDs and across the GTPases, a single model could serve as a structural model for every possible combination of GTPase and GTD, informing mutation studies and guiding drug development to disrupt the binding. Unfortunately, efforts to obtain co-crystals of any one of these complexes have been unsuccessful to date. Protein crosslinking coupled with mass spectrometry (MS) offers an alternate approach to determining residues important in these complexes.

TcdA-GTD used in crosslinking experiments was purified using published methods, using *Bacillus megaterium* as a host and expression vector pBL500[53]. The GTPases were recombinantly expressed as N-terminal glutathione S-transferase (GST) fusion proteins, using *Escherichia coli* (BL21-Star) and vectors pBL435 (RhoA), pBL436 (Rac1), and pBL437 (Cdc42). After lysis and centrifugation, cell lysates were incubated with glutathione resin (GE Healthcare, or homemade) for up to 1 h at 4 °C. Several column volumes (CV) of buffer containing 20 mM reduced glutathione were used to elute the fusion proteins. Initial experiments were performed with an N-terminal GST-tagged Rac1 protein. GST is prone to dimerization at low concentrations (~5 μM) and can confound crosslinking biochemistry[54]. To avoid the added complexity of a third protein, the GST fusion proteins were cleaved using bovine thrombin and purified over glutathione sepharose. Free Rac1 regularly appears as a doublet by SDS-PAGE and Western blot, with an apparent mass difference of 3-4 kDa. GTPases were separated

from the GST purification tag by overnight incubation with 1:100 mass ratios of bovine thrombin to GST-GTPase fusion proteins. After cleavage, the supernatant was applied to 200 µL (400 µL slurry) of benzamidine sepharose (GE Healthcare) to bind free thrombin. The flowthrough from the benzamidine column was again applied to glutathione sepharose to remove free GST protein. Flowthrough from this glutathione resin was further purified by size-exclusion chromatography using a 24 mL column packed with S75 resin (GE). Samples of pure GTPase were flash-frozen in liquid nitrogen and stored at -80 °C for future use.

Initially, we expressed GTDs with incorporated photoactivated crosslinkers using an Amber codon suppression system[55–57]. This system uses a plasmid encoding an artificial aminoacyl-tRNA synthetase (aa-RS) and tRNA which translates the canonical stop codon TAG. We used site-directed mutagenesis to alter several amino acid positions in TcdB-GTD to TAG including 309, 341, 354, and 463. These modified expression plasmids were then co-transformed with the pEVOL plasmids, which code for the tRNA and aa-RS to translate the mRNA codon UAG to the unnatural amino acids. Cultures were grown to an OD_{600} = 0.6 and induced with IPTG. Simultaneous with IPTG induction, the unnatural amino acid—*p*-arylazide phenylalanine (pAzF) or *p*-benzophenone (pBpF)—was added to the growth media for final concentration of 10 mM. While purification and UV crosslinking successful, the overall yields were quite low. This is due to the inherent inefficiencies of codon suppression, where not all UAG are properly translated by the artificial tRNA and result in premature termination. In our case, TcdB-GTD had an N-terminal 6x His-tag, which caused prematurely truncated proteins (eg., TcdB-GTD 1-306, or 1-447) to co-purify with full-length proteins. Purified

proteins with pAzF or pBpF incorporated were incubated with Rac1 on ice for up to 30 min. The samples were then exposed to 120 s of UV light (254 nm for pAzF; 345 nm for pBpF) on ice then subjected to SDS-PAGE. pAzF-labeled TcdB-GTD at positions 309, 341, and 354 was able to crosslink with GST-Rac1 and free Rac1. Additional crosslinking between TcdB 351pAzF and GST-RhoA and GST-Cdc42 was observed (Figure 2-3)

Aside from incorporating unnatural amino acids, I also explored the use of multi-functional and cleavable crosslinking agents. These small molecules included DC4 (Lys-reactive and cleavable) and sulfo-SBED (Lys reactive, cleavable, biotin labeled). Given the high concentrations of proteins employed in these experiments, it is essential that sources of crosslink noise are established and mitigated. Incubating only one protein with crosslinker was performed to establish the potential for GTD-GTD or Rac1-Rac1 background with DC-4.

Because of its nonspecific (lysine) reactivity, crosslinking TcdA-GTD and Rac1 with DC4 leads to a laddering of protein bands by SDS-PAGE (Figure 2-3). Some of these bands match with the expected molecular weight of the linked complex, but it is difficult to interpret by PAGE alone. I repeated the crosslinking experiment and performed a Western blot using an antibody against TcdB-GTD (Figure 2-4). Despite the covalent dimer of TcdB-GTD, there is a weak band observed around 80 kDa, which is consistent with crosslinking to Rac1. Incubating the protein mixture at 4 °C may stabilize the complex and increase overall yield. Similarly, use of catalytically inactive GTD (e.g., W520A, D270N) or T35/37A GTPases in combination with a glycosylation inhibitor could promote more complex crosslinking.

While SBED is an excellent compound for lysine-labelling, there are dozens of lysines on TcdA/TcdB-GTDs and GTPases. A paper reporting the synthesis of SBED also included a scheme to synthesize a similar compound—SCA1—which substitutes the lysine-reactive N-hydroxysuccinimide for a 2-pyridyldithiol moiety[58,59]. This enables the SCA1 to form a reducible bond with cysteine residues, while retaining the photoactive linker and biotin purification handle. The specificity of cysteine labelling is preferred—especially in cases where proteins have few or no native cysteines. Site-directed mutagenesis to include single cysteines putative binding sites enables significant simplification of downstream HPLC-MS analysis. Having the bait protein with a single, specific label significantly reduces the number of possible crosslinked peptides.

While SCA1 compound is not commercially available, it may be synthesized from either SBED itself, or a handful of precursors. Based on the published protocol, I attempted to produce SCA1 from SBED using 2,2'-dithiodipyridine. Analyzing the solution after incubating the reaction at room temperature for an hour, SCA1 was identified by LC-MS. Using the HPLC data, the approximate yield was 30% of starting material (1 mg). Given the expense of the SBED as a precursor, it is clearly impractical to synthesize SCA1 in this manner. Other cysteine-reactive trifunctional crosslinkers use a methanethiosulfonate in place of pyridyldithiol and are commercially available from Santa Cruz. While creating the point mutants required for the amber codon-based unnatural amino acid incorporation, I also produced C395A TcdB-GTD. Using this cysteine-free background, I produced a series of cysteine mutations (S306C, K380C, M447C, A496C and A525C) in anticipation of using a cysteine-reactive probe such as

SCA1. The affinity-purification biotin tags of SBED and SCA1 make them ideal for improving the signal in downstream LC-MS applications. However, their high cost is a limiting factor in instances where significant optimization is required. Therefore, in future explorations of crosslinking to probe the binding surfaces of these proteins, it is more fiscally practical to use bifunctional crosslinkers such as 4-methylbenzophenone (MPD). Ultimately, the success of the crosslinking experiment is dependent on the MS data—especially in the analysis of fragmentation products.

Analyzing the distribution of amino acid variation on the structure of TcdB, I found frequent changes on the concave face within of the GTD. I found that the $TcdB_{var}$ GTD had several similarities with TcsL-GTD in this region, which suggests differences within helices 15, 16, and 17 of the GTD impacts GTPase specificity. The crosslinking experiments demonstrate the feasibility of this technique in future studies of GTD and GTPase interactions. The yields of the crosslinked complex are too low for crystallographic studies, but are more than sufficient for MS analysis. Unnatural amino acid incorporation was inefficient, but demonstrated successful crosslinking between TcdB-GTD pAzF at position 309, 341, 354, 463 and Rac1. Crosslinks between TcdB-GTD with pAzF at position 341 was also observed with Cdc42 and RhoA. Yields of the crosslinked complex did not significantly change when UDP-glucose or UDP were present, which suggests that the presence of a ligand does not significantly alter the concentration of complex at equilibrium.

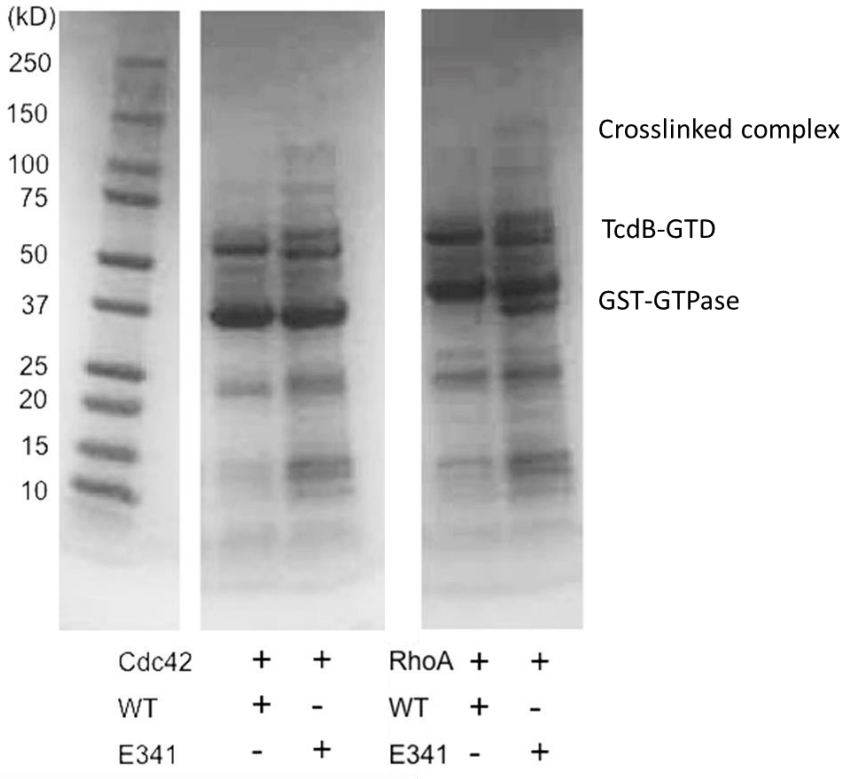

Figure 2-3. Photoactivated crosslinking between TcdB-GTD (341-pAzF) and GST-GTPase fusion proteins. Each reaction contained 1 μM of GST-GTPase and TcdB-GTD wildtype (WT) or the E341pAzF labeled TcdB-GTD. Samples were exposed to 254 nm UV light for 120 seconds.

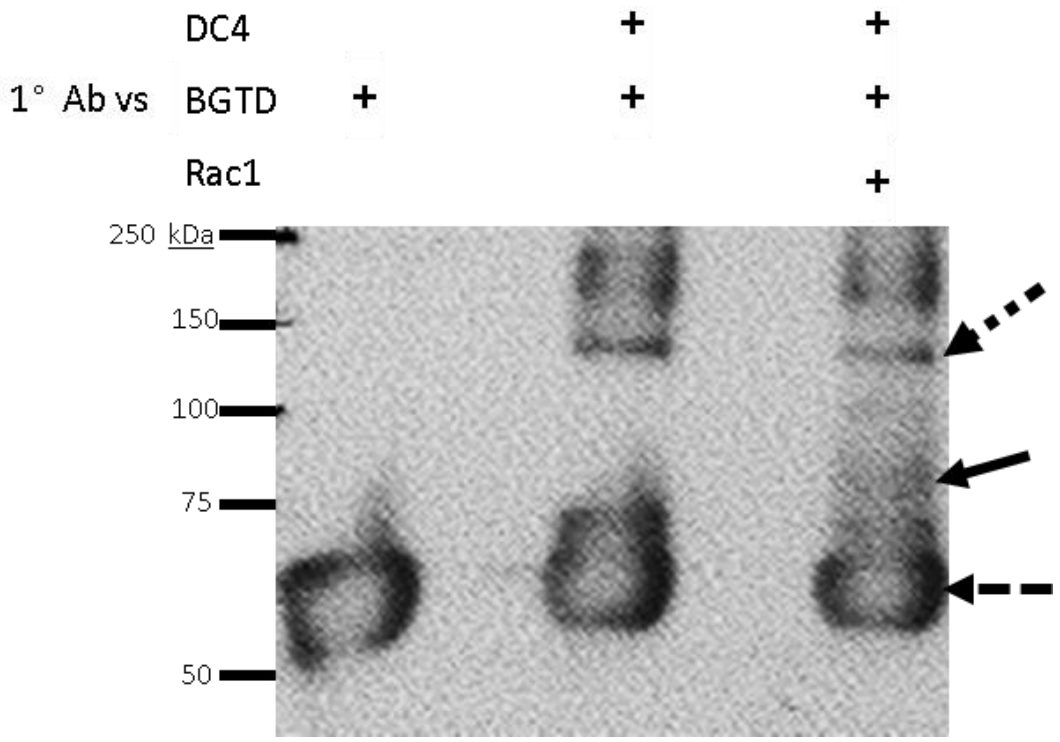

Figure 2-4. TcdB-GTD and Rac1 crosslinked with DC4. The dashed arrow indicates TcdB-GTD (64 kDa), the dotted arrow indicates a crosslinked dimer of TcdB-GTD (~128 kDa), and the solid arrow represents Rac1 crosslinked to the GTD (~88 kDa). The PVDF membrane was probed with a monoclonal (20B3) mouse anti-TcdB-GTD antibody.

Chapter 3

Clostridium difficile toxin glucosyltransferase domains in complex with a non-hydrolyzable UDP-glucose analogue

3-1 Introduction

C. difficile is the leading cause of hospital-acquired diarrhea and pseudomembranous colitis worldwide. The organism produces two homologous toxins, TcdA and TcdB, which enter and disrupt host cell function by glucosylating and thereby inactivating key signalling molecules within the host. As a toxin-mediated disease, there has been a significant interest in identifying small molecule inhibitors of the toxins' glucosyltransferase activities. This study was initiated as part of an effort to identify the mode of inhibition for a small molecule inhibitor of glucosyltransferase activity called apigenin[60]. In the course of trying to get co-crystals with this inhibitor, we determined five different structures of the TcdA and TcdB glucosyltransferase domains and made use of a non-hydrolyzable UDP-glucose substrate. While we were able to visualize apigenin bound in one of our structures, the site was a crystal packing interface and not likely to explain the mode of inhibition. Nevertheless, the structure allowed us to capture an apo-state (one without the sugar nucleotide substrate) of the TcdB

glycosyltransferase domain that had not been previously observed. Comparison of this structure with structures obtained in the presence of a non-hydrolyzable UDP-glucose analogue have allowed us to document multiple conformations of a C-terminal loop important for catalysis. We present our analysis of these five new structures with the hope that it will advance inhibitor design efforts for this important class of biological toxins. *Clostridium difficile* is a spore-forming anaerobe that produces two large, homologous toxins. The toxins, TcdA and TcdB, are the primary virulence factors for *C. difficile* infection (CDI) and are part of the large clostridial toxin (LCT) family. Members of the LCT family share sequence homology, domain organization, and common origins[20,34,61]. In addition to TcdA and TcdB, the LCTs include virulence factors produced by the pathogens C. sordellii (TcsL and TcsH), C. novyi (Tcnα), and C. perfringens (TpeL)[4,30,62]. Along with the primary sequence homology of the holotoxins, the published GTD structures of TcdA, TcdB, TcsL, and Tcnα reveal that the LCT-GTDs also share structural homology[24,53]. The GTDs can be organized into four domains, which includes a membrane localization domain (MLD) (Figure 3-1, yellow), the glycosyltransferase-A fold (Figure 3-1, blue), a globular subdomain (Figure 3-1, orange), and two helical clusters (Figure 3-1, green). An important element within LCT-GTDs is a conserved tryptophan, which resides on a flexible helix and plays a role in catalysis (Figure 3-1, inset). Previous work has shown that this tryptophan plays a significant role in glycosyltransfer. In TcdB-GTD W520A, glucosyltransferase k_{cat} is reduced over 800-fold, while the UDP-glucose K_m varies only slightly[52]. This study demonstrated the importance of this conserved tryptophan in catalysis of

glucosyltransfer, but not in UDP-glucose binding. The first structures of TcdB-GTD were obtained by including UDP-glucose and cofactor Mn^{2+} in the crystallization conditions[63]. The electron density maps revealed TcdB bound to UDP and glucose, indicating that the substrate had been hydrolyzed. In contrast, no hydrolysis was seen in initial structures emerging from co-crystallization of the TcdA GTD with UDP-glucose and Mn^{2+} cofactor[53]. These observations are consistent with kinetic data indicating that in the absence of target GTPases, the GTDs hydrolyze UDP-glucose with TcdB-GTD having a five-fold higher Vmax compared to TcdA-GTD[64,65].

To date, only two GTDs from the LCT family have crystal structures under apo conditions: TcdA-GTD and Tcnα-GTD. To better understand the structural changes associated with substrate binding, we set out to investigate, 1) an intact substrate bound to TcdB-GTD, 2) an apo form of TcdB-GTD, and 3) the effects and/or binding sites of the small molecule apigenin. In this study, we present crystal structures of TcdA and TcdB-GTDs in complex with UDP-2-deoxy-2-fluoroglucose (U2F), a non-hydrolyzable UDP-glucose analogue. Additionally, we found that apigenin functioned to bridge two TcdB-GTD chains, giving rise to a new crystal form and space group. These conditions, which required the presence of apigenin, enabled the crystallization of TcdB-GTD in an apo-like form. Together, these new structures provide insight into the flexibility within the *C. difficile* toxin GTD active sites.

Each LCT contains four domains: a glycosyltransferase domain (GTD), autoprotease domain (APD), delivery domain, and combined repetitive oligopeptides (CROPs) domain. The toxins bind and enter host cells through receptor-mediated endocytosis. During endosomal acidification, a conformational change in the delivery

domain facilitates translocation of the GTD and APD across the endosomal membrane into the cell cytosol[66]. The APD is activated by IP6 and cleaves the GTD at its C-terminus, thus releasing the GTD into the cell[67]. The LCT-GTDs catalyze the transfer of a sugar from uridine diphosphate (UDP) to a regulatory domain of host cell GTPases[68]. The TcdA and TcdB GTDs target Rho family GTPases—including RhoA, Rac1, and Cdc42[69]. These GTPases are essential regulators of focal adhesions, actin organization, cell morphology, and migration. Glycosylation by TcdA and TcdB GTDs leads to loss of focal adhesions, F-actin depolymerization, and apoptotic cell death [22,24,70].

Along with the primary sequence homology of the holotoxins, the GTD structures of TcdA, TcdB, TcsL, and Tcnα reveal that the LCT-GTDs also share structural homology[22,53,63,71]. The GTDs can be organized into four domains, which includes a membrane localization domain (MLD) (Figure 3-1, yellow), the glycosyltransferase-A fold (Figure 3-1, blue), a globular subdomain (Figure 3-1, orange), and two helical clusters (Figure3-1, green). An important element within LCT-GTDs is a conserved tryptophan, which resides on a flexible loop at the GTD C-terminus, with proximity to UDP-glucose (Figure 3-1, inset, magenta). Mutation of this tryptophan affects catalysis, but not UDP-glucose binding. Specifically, in TcdB-GTD W520A, the k_{cat} of glucosyltransfer is reduced over 800-fold compared to wildtype, while the UDP-glucose K_m varies only slightly[52]. The first structures of TcdB-GTD were obtained by including UDP-glucose and cofactor Mn^{2+} in the crystallization conditions[63]. The electron density maps revealed TcdB bound to UDP and glucose, indicating that the substrate had been hydrolyzed. In contrast, no hydrolysis was seen in structures emerging from co-

crystallization of the TcdA GTD with UDP-glucose and Mn^{2+} cofactor[53]. These observations are consistent with kinetic data indicating that in the absence of target GTPases, TcdB-GTD will hydrolyze UDP-glucose with a five-fold higher V_{max} compared to TcdA-GTD[64,65]. To date, only two GTDs from the LCT family have crystal structures under apo conditions: TcdA-GTD and Tcnα-GTD[22,53].

To better understand the structural changes associated with substrate binding, we set out to investigate, 1) an intact substrate bound to TcdB-GTD, 2) an apo form of TcdB-GTD, and 3) the effects and/or binding sites of the small molecule inhibitor apigenin. In this study, we present crystal structures of TcdA and TcdB-GTDs in complex with UDP-2-deoxy-2-fluoroglucose (U2F), a non-hydrolyzable UDP-glucose analogue. Additionally, we show that apigenin can bridge two TcdB-GTD chains, giving rise to a new crystal form and space group, one that allows visualization of TcdB-GTD in an apo-like form. Together, these new structures provide insight into the range of flexibility associated with the catalytic tryptophan loop when moving from *apo* to UDP-glucose bound conditions.

3-2 Methods

3-2-1 Synthesis of UDP-2-deoxy-2-fluoroglucose

U2F was synthesized through Vanderbilt's Small Molecule Synthesis Core according to published methods[72]. Sample purity was determined by liquid chromatography, mass spectrometry, and the final structure validated by $^{1}H/^{13}C$ nuclear magnetic resonance. Samples of dry U2F were resuspended in 20 mM HEPES pH 8, 50 mM NaCl and stored at -20 °C.

3-2-2 Purification and Crystallization of TcdA-GTD

TcdA-GTD was purified according to published methods using the *Bacillus megaterium* expression system plasmid pCHis1622 (MoBiTec, Germany)[53]. Briefly, *B. megaterium* overnight cultures at 37 °C were diluted into fresh LB-Miller broth. At OD_{600} = 0.3, protein expression was induced with 5 g/L xylose, and the temperature was reduced to 18 °C. After 16-18 h, cultures were pelleted and frozen at -80 °C. After thawing, cell pellets were lysed using an Emulsiflex C3 (Avestin Inc., Ottawa, Canada) at 20,000 lb/in^2 and centrifuged at 16,000 x g for 30 min at 4 °C to remove membranes and insoluble components. Lysis buffer was composed of 50 mM HEPES pH 8, 500 mM NaCl, and 2 mM imidazole. Approximately 10 µg of DNAse-I (Sigma-Aldrich) and final concentrations of 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM Leupeptin, and 1 mM Pepstatin were added to lysis buffer. The lysate was incubated with Co^{2+}-charged chelating resin (GE Healthcare) for 30 min at 4 °C and washed with three column volumes of lysis buffer. Purified TcdA-GTD was eluted with 20 mM HEPES pH 8, 50 mM NaCl and 300 mM imidazole. Elutions were concentrated and purified using a 5 ml HiTrapQ anion exchange column (GE Healthcare). Elutions containing TcdA-GTD were concentrated and further purified over a 24 mL S-200 Sepharose column (GE Healthcare). Pure TcdA-GTD aliquots in 20 mM HEPES pH 8, 150 mM NaCl were flash-frozen with liquid N_2 and stored at -80 °C. All proteins were frozen at concentrations ≥ 10 mg/mL (~150 µM for TcdA-GTD).

Prior to crystallization, aliquots were rapidly thawed, subjected to size-exclusion chromatography (SEC), and concentrated using Amicon Ultra centrifugal filters with a 50 kDa cutoff. For co-crystals with UDP-glucose or U2F, a final concentration of 1 mM

of the UDP-conjugate was used. All protein solutions also included 2-10 mM $MnCl_2$. All TcdA-GTD co-crystals were obtained by hanging drop diffusion with a 1:1 ratio of 5-12 mg/mL protein solution to mother liquor comprised of 0.1 M HEPES pH 7-8.5, 0.2 M L-proline, and PEG 3350 10-25%. DMSO stocks of 0.1 M apigenin were diluted into concentrated protein solutions to obtain final concentrations of 0.5-10 mM. We observed significant compound precipitation at these concentrations; however, the total protein concentration was not affected. Large, hexagonal needles grew within 24-48 h at 21 °C. Crystals were exchanged into mother liquor containing 15-20% glycerol, (including 0.5-10 mM apigenin where indicated) mounted on cryo-loops and flash-frozen in liquid N_2.

3-2-3 Purification and Crystallization of TcdB-GTD

TcdB-GTD was purified similar to previous methods[53]. The sequence corresponding to TcdB-GTD_{1-543} was copied from the parent plasmid pBL149 and cloned into pET8a+ using restriction sites BamHI 5' and XhoI 3' to yield pBL720. Briefly, *E. coli* (BL21-Star) glycerol stocks were grown in LB-Miller broth at 37 °C. At OD_{600} = 0.6, protein expression was induced with 1 mM IPTG, and the temperature was reduced to 18 °C. After 16-18 h, cultures were pelleted and frozen. After thawing, cell pellets were lysed using an Emulsiflex C3 at 18,000 psi and centrifuged at 16,000 x g to remove membranes and insoluble components. Lysis buffer was composed of 50 mM HEPES pH 8, 500 mM NaCl, and 2 mM imidazole. Approximately 10 µg of DNAse-I (Sigma-Aldrich), 1 mM PMSF, 1 mM Leupeptin, and 1 mM Pepstatin were added to lysis buffer. The lysate was purified by successive immobilized metal affinity, anion exchange, and SEC.

Aliquots of concentrated protein were frozen in liquid N_2 and stored at -80 °C. For crystallization, aliquots were thawed, purified using SEC, and re-concentrated. Apigenin was dissolved in 100% DMSO at a stock concentration of 0.1 M. Aliquots were stored at -20 °C until used. The DMSO apigenin solution was added to concentrated protein to obtain final concentrations of 0.5-10 mM. We observed significant compound precipitation at these concentrations; however, the total protein concentration was not affected. TcdB-GTD crystals with apigenin were obtained by hanging drop diffusion with a 1:1 ratio of 8-15 mg/mL protein solution (20 mM HEPES pH 8, 150 mM NaCl, 1-10 $MnCl_2$) to mother liquor comprised of 0.1 M HEPES pH 7.5-8.5, 0.2 M $Mg(CH_3CHOO)_2$, and PEG 3350 5-25%. Rhombohedral crystals grew within 24-48 h only in the presence of apigenin. Crystals were exchanged into mother liquor containing 10-20% glycerol, (0.5-10 mM apigenin where indicated) mounted on cryo-loops and flash-frozen in liquid N_2. TcdB-GTD + U2F co-crystals were obtained by hanging-drop diffusion using conditions previously reported with a mother liquor of 0.1 M MES pH 6-6.5, 0.2 $(NH_4)_2SO_4$, and PEG 8K 16-34% [63]. This was mixed 1:1 with TcdB-GTD at 10-15 g/L in 20 mM HEPES pH 8, 150 mM NaCl, 1 mM U2F, and 1-10 mM $MnCl_2$. Short bars and biaxial clover crystals formed within 72 h at 21 °C. Crystals were transferred into mother liquor containing 15-20% glycerol, including 0.5-10 mM apigenin (where indicated) mounted on cryo-loops and flash-frozen in liquid N_2.

Table 3-1. Crystallography statistics.

Protein complex	TcdB-GTD U2F 5UQN	TcdA-GTD U2F 5UQL	TcdB-GTD U2F + apigenin 5UQM	TcdA-GTD U2F + apigenin 5UQK	TcdB-GTD apigenin 5UQT
Data processing	xia2	HKL2000	xia2	HKL2000	xia2
Space group	$P4_12_12$	$P6_5$	$P4_12_12$	$P6_5$	$C222_1$
Cell dimensions					
a, b, c (Å)	62.39, 62.39, 328	142.1, 142.1, 63	61.9, 61.9, 325.98	142.8, 142.8, 66.1	82.09, 154.81, 271.4
α, β, γ (degrees)	90, 90, 90	90, 90, 120	90, 90, 90	90, 90, 120	90, 90, 90
Resolution (Å)	54.19-2.06	46.5-1.97	42.28-2.03	50.00-1.85	44.44-2.75
Rmerge					
I/σI	23.31 (3.75)	21.35 (1.55)	23.89 (2.91)	28.4 (1.32)	21.33 (1.4)
Completeness (%)	97.84 (93.26)	99.9 (99.9)	99.6 (92.9)	99.6 (100.0)	99.75 (99.93)
Redundancy	11.3 (10.3)	8.2 (7)	11.3 (8.7)	7.4 (6.8)	6.6 (6.8)
Wavelength (Å)	0.97856	0.97872	0.97856	0.97856	0.97856
Unique observations	41228	51462	42362	65449	45388
CC1/2	0.946	0.855	0.848	0.865	0.549
Refinement					
Resolution (Å)	54.19-2.06	46.5-1.97	42.2-2.03	31.4-1.85	44.4-2.75
No. of reflections	41070	51428	42208	65417	45276
Rwork/Rfree	0.236/0.203	0.204/0.184	0.228/0.197	0.207/0.182	0.236/0.196
No. of atoms (non H)	4789	5002	4768	5058	8842
Avg. protein B-factor ($Å^2$)	43.54	39.63	42.69	52.58	102.66
RMSD values					
Bond lengths (Å)	0.008	0.004	0.014	0.021	0.007
Angles (degrees)	1.019	0.665	1.264	1.598	0.906
Ramachandran plot (%)					
Most favored	97.4	98.52	97.59	97.01	96.83
Allowed	2.41	1.48	2.41	2.99	3.08
Disallowed	0.19	0	0	0	0.09

3-2-4 Data collection and refinement

Data were collected from single crystals on LS-CAT beamlines 24 ID-D, F, and G at the Advanced Photon Source (Argonne National Laboratory, IL) at 100 K. Data were indexed and scaled using HKL2000 or the program suite xia2[73,74]. Within xia2, XSCALE and XDS were used to index with peaks from all images[73]. TcdA-GTD data collected from crystals without UDP substrates were phased using PHENIX-Phaser-MR with coordinates from Protein Data Bank (PDB) 3SS1, while substrate-bound crystals were phased with PDB 3SRZ. TcdB-GTD data were phased using PHENIX-Phaser-MR with coordinates from PDB 2BVL[53,63,75]. Models were iteratively built with Coot and refined via PHENIX with 4-5 TLS groups per chain[75,76]. Final coordinates for TcdA-GTD co-crystallized with U2F and apigenin (5UQK) and U2F alone (5UQL) were deposited into the PDB. In TcdB-GTD datasets indexed as C2221, two chains exhibiting non-crystallographic symmetry (NCS) were identified in the asymmetric unit (ASU). Alternatively, crystals grown under previously reported conditions were indexed as P41212 and contained a single chain in the ASU. NCS torsion-angle restraints were used in refining the C2221 structure. RMSD values were calculated and figures were generated using PyMol[77]. Final coordinates for TcdB-GTD co-crystallized with U2F (5UQN), U2F and apigenin (5UQM), and apigenin alone (5UQL) were deposited into the PDB.

3-3 Results

3-3-1 TcdA-GTD and TcdB-GTD in complex with U2F

UDP-2-deoxy-2-fluoroglucose (U2F) and related UDP-fluoro-sugars have been used to crystallize a number of diverse glycosyltransferases[72,78–80]. The U2F as a substrate renders hydrolysis or transfer by glycosyltransferases unfavorable. The GTDs of TcdA and TcdB were crystallized in the presence of U2F, and structures were determined at 2.0 and 2.06 Å, respectively (Table 3-1). Published structures of the GTDs bound to UDP-glucose were used as models for molecular replacement (3SRZ for TcdA and 2BVL for TcdB), and the refined U2F structures aligned to their starting models with C_α RMSD values of 0.25 $Å^2$ (TcdA) and 0.44 $Å^2$ (TcdB).

The electron density clearly shows the presence of U2F within the GTD active sites of TcdA (Figure 3-2A) and TcdB (Figure 3-2B). The TcdB-GTD + U2F structure gave us an opportunity to visualize any changes from TcdB-GTD bound to hydrolyzed UDP and glucose. The primary changes between an intact substrate are found in the position of glucose when still bound to UDP or after hydrolysis. We found that while the glucose ring tilts slightly, the residues which interact with glucose do not significantly change upon substrate hydrolysis (Figure 3-3). Overall, active site residues of both TcdA and TcdB-GTD bound to U2F closely match the UDP-glucose structure of TcdA and the UDP + glucose structure of TcdB. Comparing TcdA-GTD+U2F to the structure co-crystallized with UDP-glucose, we observe nearly identical positions of residues in the active site.

The UDP moiety is held in place by a combination of hydrogen bonds, aromatic interactions, and a Mn^{2+} cofactor (Shown for TcdB in Figure 3-4, but also true in TcdA). A conserved tryptophan (TcdA W101, TcdB W102) anchors the uracil ring with aromatic

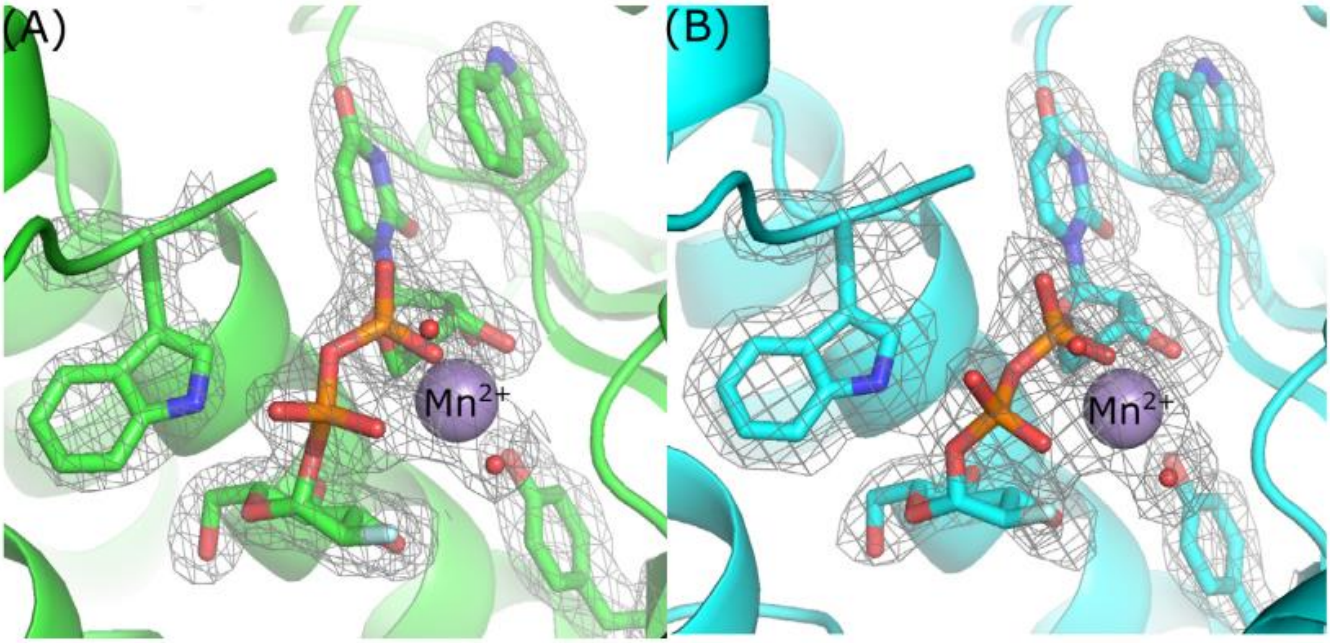

Figure 3-2. 2mFo-DFc density map contoured at 1.0 sigma of select aromatic groups and U2F in the active site of TcdA-GTD **(A)** and TcdB-GTD **(B)**[60].

π-stacking (Figure 3-4). In the context of TcsL, the analogous TcsL-GTD W102A mutant has a 2000-fold higher K_m and a 16-fold lower k_{cat} compared to wild-type TcsL-GTD[81]. The W102Y TcsL-GTD mutant was slightly more active than the W102A mutant with only a 21-fold increase in K_m and four-fold reduction in k_{cat}[81]. This suggests that the aromatic nature of the tryptophan is significant for interaction with the uracil ring. The fact that W102A TcdB-GTD also has significantly impaired glycosyltransferase and hydrolase activities suggests that this role will be common across the LCT family of GTDs[81,82].

The U2F of each structure interacts with other conserved elements within the GTD active sites. The uracil C2 and C4 carbonyls form hydrogen bonds with the I101 backbone amine and N139 amide nitrogens, respectively (shown for TcdB, Figure 3-4). The backbone carbonyl of I101 forms a third hydrogen bond with uracil N3, and the V287 backbone amine forms a hydrogen bond with the ribose 3'-hydroxyl. The ribose 2'-hydroxyl interacts via hydrogen bonds to the V103 backbone carbonyl, Y284 sidechain hydroxyl group, and S269 sidechain hydroxyl group (Figure 3-4B). In TcdB-GTD, it was demonstrated that Y284A constructs have over 1000-fold lower glucosyltransferase activity as compared to wild-type. This interaction is also observed in crystal structures of TcdA-GTD, which suggests Y283 plays a similarly important role by coordinating with the ribose 2'-hydroxyl in TcdA [52,53].

A key element in glycosyltransferases is the conserved DxD motif (D286 and D288 in TcdB), where two aspartates bind Mn^{2+}, which in turn supports the interactions with the nucleotide diphosphates (Figure 3-4B). Mutation of both DxD aspartates to asparagines in TcdA-GTD and TcsL-GTD reduces glycosyltransfer by three orders of

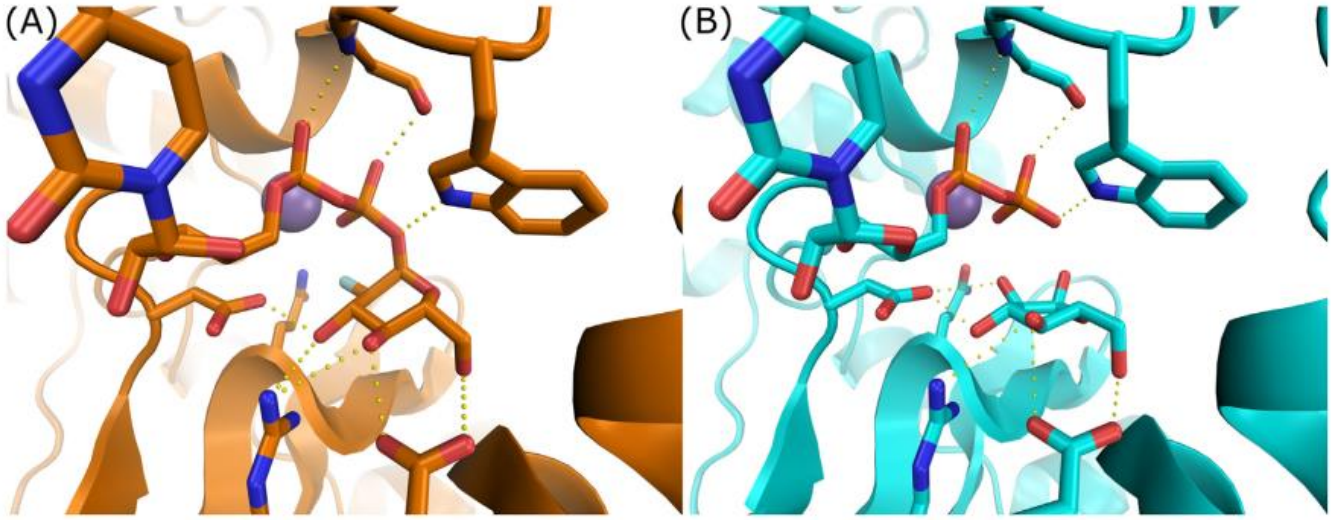

Figure 3-3. Comparison between U2F **(A)** and UDP + glucose **(B)** co-crystal structures of TcdB-GTD[60].

magnitude, demonstrating the essential role of this network of interactions[69,83]. The octahedral Mn^{2+} coordination system is completed by the diphosphate backbone of the substrate and two additional water molecules (Figure 3-4A). Typically, retaining glycosyltransferases have aspartates from the DxD motif directly interact with the metal cofactor[84]. However, we note that within the LCT family, the Mn^{2+} is directly coordinated by D288 and E515 while D286 coordinates Mn^{2+} indirectly through a water molecule. This LCT architecture is similar to the galactosyltransferase LgtC from *Neisseria meningitidis*[78].

3-3-2 Apigenin and the crystallization of TcdB-GTD in an apo-like conformation

Tam et al. recently reported that select flavonoids such as phloretin and apigenin inhibit cell rounding caused by TcdA and TcdB[39]. *In vitro* tests with purified TcdA and TcdB-GTD indicated that these flavonoids were significantly more potent in inhibiting glucosyltransferase compared to hydrolytic activities. The inhibition was independent of UDP-glucose concentration, suggesting that phloretin and apigenin function as noncompetitive inhibitors of the GTDs. Apigenin is approximately five-fold more effective at inhibiting the glucosyltransferase activity and has comparable solubility with phloretin. To better understand the effects and mechanism of action for these inhibitors, we sought to identify the apigenin binding site(s) in TcdA and TcdB-GTD crystal structures. We performed a series of screens and crystal soaking experiments with apigenin. These included multi-factorial broad screens as well as the replication of previously reported crystallization conditions. Since there was no information as to which

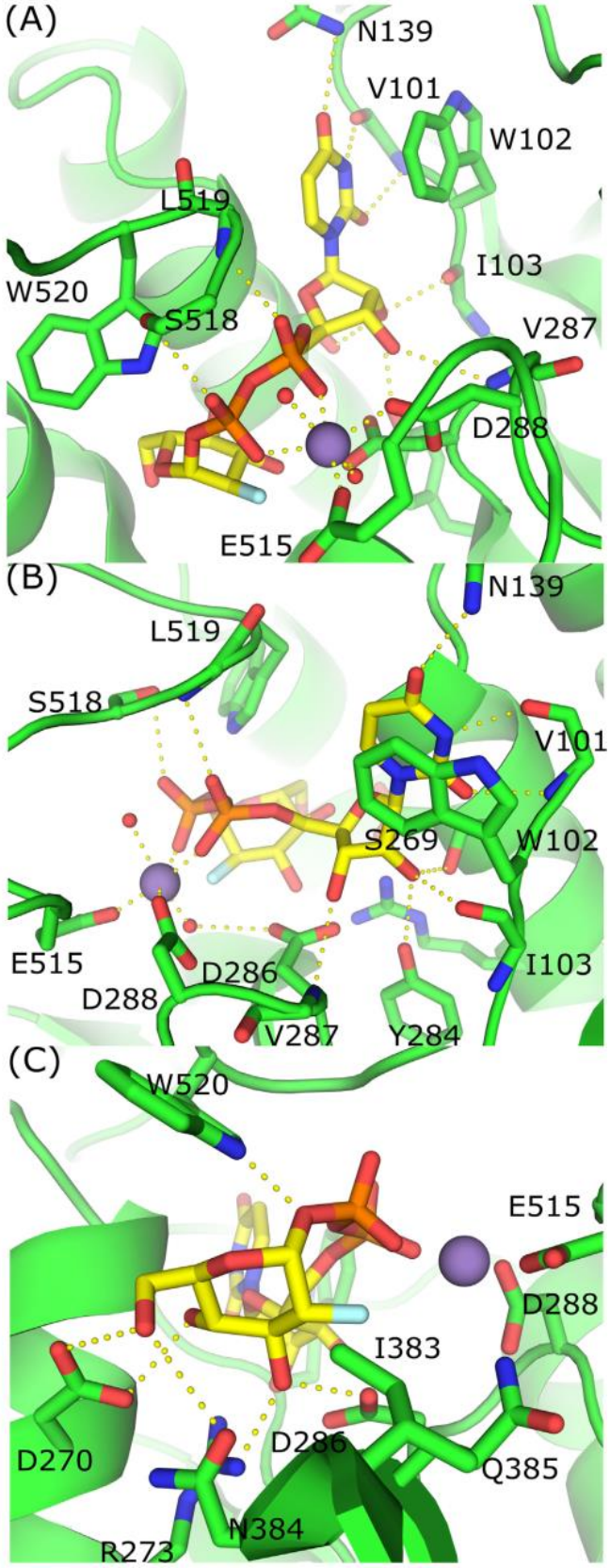

Figure 3-4. Coordination of UDP-2-deoxy-2-fluoroglucose by TcdB-GTD. **(A)** π-stacking and hydrogen bonding between UDP, Mn^{2+} and TcdB-GTD. **(B)** Alternate view of interactions between active site residues and UDP and Mn^{2+}. **(C)** Interactions between active site residues and 2-deoxy-2-fluoroglucose[60].

conformation apigenin would bind, we set up *apo* conditions along with trials that contained U2F. This allowed sampling of conditions with TcdB-GTD bound to a substrate, without the complication of hydrolysis.

We obtained structures of TcdA and TcdB-GTD in complex with U2F grown in the presence of apigenin (Table 3-1). For TcdB-GTD, no differences were observed between the structure containing U2F and the crystal grown in the presence of U2F and apigenin. We observed high occupancy of U2F within the active site. This is consistent with apigenin inhibiting glucosyltransferase activity without competing for UDP-glucose binding. However, there was no unaccounted density which would represent one or more copies of apigenin within the structure.

Despite success in obtaining a high-resolution dataset (1.85 Å), we again did not observe apigenin in the TcdA-GTD + U2F + apigenin structure. However, we observed an *apo* conformation of the W519 loop, despite the presence of U2F in the active site (Figure 3-5, orange). This everted *apo* conformation was also observed in the recent $TcdA_{1\text{-}1832}$ crystal structure[12]. We considered that apigenin may interact with W519 in TcdA-GTD in a manner that stabilized this *apo* conformation. We analyzed Polder maps (a difference map omitting bulk solvent) to scan small areas (up to five contiguous residues) with moderate to high solvent exposure, including the W519 loop. While we did not locate apigenin, the Polder maps indicated positive density similar to an intermediate position of W519 (Figure 3-5, cyan) from TcdA-GTD bound to UDP[71,85].

As previously discussed, earlier structures of TcdA have indicated that the conformation of the W519 loop depends on the presence or absence of UDP or UDP-

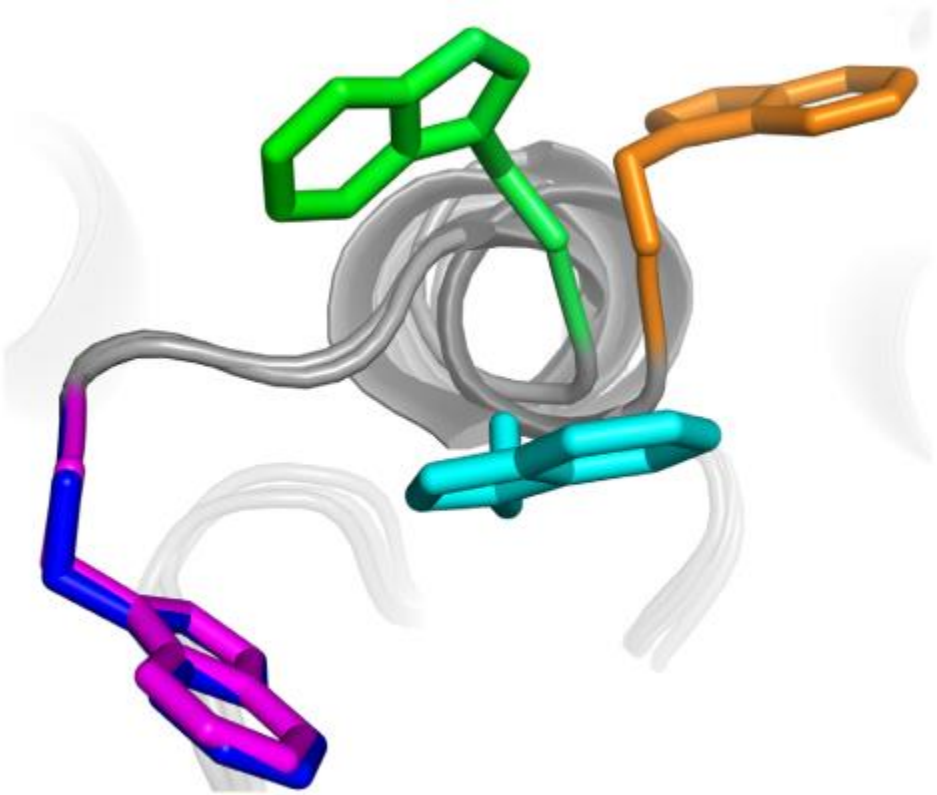

Figure 3-5. Superimposed structures of W519 in TcdA-GTD with UDP-glucose (blue), apo TcdA-GTD (green), TcdA-GTD with UDP (cyan), and apo TcdA 1–1832 (orange). TcdA-GTD bound to U2F (magenta) or UDP-glucose have nearly identical W519 conformations[60].

glucose within the active site. TcdB-GTD had only been crystallized in a single conformation, in complex with UDP + glucose. In an effort to better characterize the conformations and flexibility of the TcdB W520 loop, we sought to crystallize TcdB-GTD in an *apo* form. Crystals of TcdB-GTD alone diffracted poorly (>7 Å) and were difficult to reproduce as single, non-twinned crystals. The inclusion of apigenin in crystal screens with TcdB-GTD yielded several new conditions and a different space group. We collected a 2.8 Å dataset from TcdB-GTD crystals grown without UDP-glucose or U2F, which depended on the presence of apigenin (Table 3-1). The data were indexed as space group $C222_1$, with a relatively large unit cell (dimensions 92 Å, 154 Å, 271 Å). Molecular replacement led to a solution where two copies of TcdB-GTD were joined along alpha helices 15 and 16 within the asymmetric unit (ASU) (Figure 3-6A). Overall, both chains within the ASU were highly similar (RMSD of 0.29 $Å^2$) with minor differences at the C-terminus and an unstructured loop bridging two helices (residues 156-171). Both chains within the TcdB-GTD ASU demonstrated *apo*-like conformations in the W520 loop, though the precise position of W520 differed (Figure 3-6B).

After some iterative refinement, we observed a consistent area of positive density in our difference maps at the junction of the two ASU chains. These differences were coincident with a planar ellipse of density in 2mFo-DFc maps at the same location that looked similar in shape and size to apigenin (Figure 3-6C). The inclusion of apigenin at this chain-chain interface explains the new crystal form and the dependence of this form on apigenin. In this instance, we propose that apigenin merely fills a hydrophobic void between the GTD chains in the ASU of our TcdB-GTD structure (Figure 3-6D).

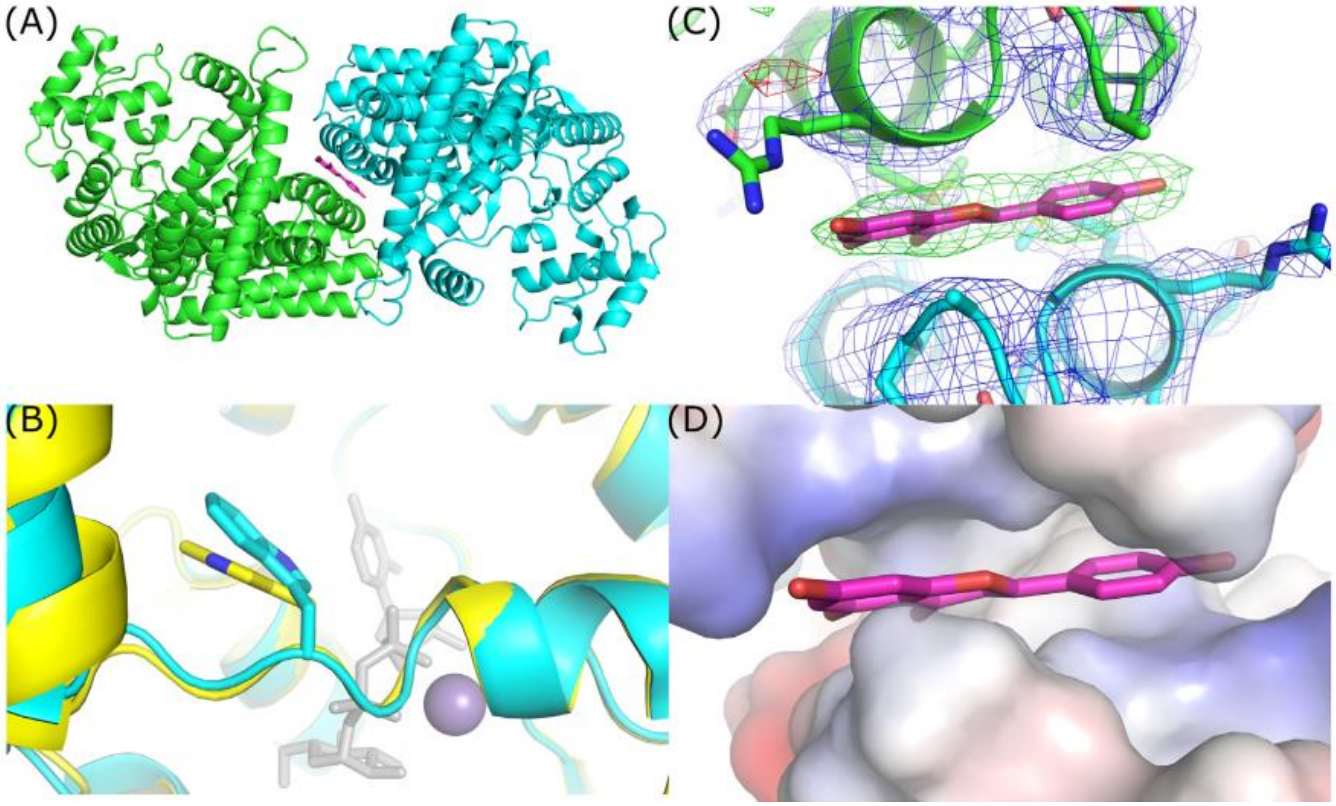

Figure 3-6. TcdB-GTD crystallized in an apo-like conformation contains two chains within the ASU. (A) Top-down view demonstrating the position of apigenin and non-crystallographic symmetry of the TcdB-GTD chains. (B) Overlay of apo TcdA-GTD (yellow) and TcdB-GTD (cyan) in an apo-like conformation. (C) 2mFo-DFc and Fo-Fc maps of apigenin site contoured at 1 and 4 σ respectively. (D) Vacuum electrostatic model of the hydrophobic patch occupied by apigenin[60].

3-4 Discussion

The growing number of deposited LCT-GTD structures allows for a more comprehensive analysis of GTD function. The primary difference between *apo* and UDP-glucose bound states across TcdB, TcsL, TcdA, and Tcnα-GTDs are the conformations of the active site "lid" that includes the conserved 519/520 tryptophan. (Figure 3-1, magenta). Crystal structures of multiple LCT-GTDs suggest that the position of the tryptophan is related to the presence or absence of substrate.

The relationship between this tryptophan loop and UDP-glucose is clear when comparing structures of TcdA-GTD in *apo* and UDP-glucose bound forms. In these cases, TcdA-GTD was crystallized in the same space group and the W519 loop was shifted into the active site in the presence of UDP-glucose[53]. TcdA-GTD bound to UDP-glucose adopted a conformation nearly identical to that of TcdB-GTD, though it contains an intact substrate (Figure 3-5, blue). TcsL-GTD was crystallized in the presence of UDP-glucose and shares the substrate bound conformation with TcdA and TcdB-GTD[22]. Tcnα-GTD, however, was determined in an *apo* state. The *apo* form of TcdA-GTD is similar to that of Tcnα-GTD, where the loop tryptophan was rotated outside of the active site (Figure 3-5, green). Two additional TcdA-GTD structures show the conserved W519 in identical conformations under *apo* and UDP-bound conditions (Figure 3-5, cyan)[71]. This conformation could represent an intermediate state between the *apo* and substrate-bound forms. Importantly, the absence of glucose in the UDP-bound structure suggests that UDP is insufficient to stabilize the loop and tryptophan within the active site. This correlates well with the conformation of W520 in TcdB-GTD in complex with

hydrolyzed UDP-glucose. In this case, the glucose was retained in the active site and likely stabilizes W520 in its inward, substrate-bound conformation.

3-5 Conclusion

In this study, we report new crystal structures of *C. difficile* toxin glucosyltransferase domains obtained using a non-hydrolyzable UDP-glucose analog, U2F, some in the presence of apigenin—a reported inhibitor of TcdA and TcdB-GTDs. We found that despite the hydroxyl to fluorine substitution in the synthetic compound, the active site and sugar selection residues remained unchanged in both proteins. While the inclusion of apigenin was required for crystallizing an *apo*-like form of TcdB-GTD, the molecule was not observed in the other crystals, and the mechanism for this inhibitor remains unclear. In determining these new structures of TcdA and TcdB-GTD we observed disparate conformations of the active site tryptophan. These conformations add to the existing group of structures where a conserved tryptophan samples a broad conformational space in the context of a comparatively static active site.

Chapter 4

Conclusions and Future Directions

4-1 Conclusions

C. difficile is the leading cause of hospital-acquired diarrhea and pseudomembranous colitis worldwide. The major virulence factors in CDI (TcdA and TcdB) are members of the LCT family. Due to the growing global healthcare burden of *C. difficile*, it remains the most-studied of the LCT-producing Clostridia. In an effort to further our understanding, I sought to study the structural characteristics of the TcdA/B-GTD using a combination of sequence comparison, structural modeling, crosslinking, and crystallography with substrates and inhibitors.

Structural work among the LCTs has revealed consistent conservation of function and motifs, especially in the GTDs. Partial (e.g., CROPs) and full domains of TcdA have been crystallized, but significant information about intra-domain contacts was revealed with the structure of $TcdA_{1\text{-}1832}$. One of the important discoveries is the conformational change induced by IP6 (Figure 1-4). An interesting aspect of TcsL is its IP_6 binding is less affected by pH compared to TcdB[86]. The hypothesis that *C. sordellii* is more common in acidic environments is supported by its association with gynecological complications. While the pathogen is rare, the characteristics providing acid-resistance are worth further investigation. While there are data supporting pore-formation by TcdB, the fundamental mechanism of GTD-APD translocation is unclear.

Based on EM of TcdA under neutral and acidic conditions (Figure 1-3), it is clear that pH change is a driving force behind membrane insertion in the LCTs. Currently, elements within the delivery domain sufficient for membrane insertion and translocation are unknown.

The recent identification of receptors for TpeL and TcdB represent significant steps toward understanding the LCTs. The TpeL receptor—LRP1—supports a model of intoxication which is independent of the CROPs present in the five other LCTs. Similarly, the CROPs domain was shown to be not necessary for TcdB to bind CSPG4, PVRL3, or FZD. There are data for demonstrating expression of RNA for CSPG4, PVRL3 and FZD1 within the colon[87,88]. Given the number of receptor candidates identified from animal models and *in vitro* studies, it is unclear which proteins drive the pathology of *C. difficile* infection. However, working towards a structural understanding of the interactions between toxin and receptor is an important direction for future studies.

Homology modeling based on TcdA and TcdB is a rational approach given the significant sequence identity across the toxins and the difficulty of working with six massive proteins. In Chapter 2, I presented a structural analysis of clinical $TcdA^-/TcdB^+$ strains, which have a non-canonical CPE and related variation in GTPase glucosylation specificity. TcdB-GTD variation is concentrated along helices proposed to determine GTPase target specificity. TcdB-GTD variants which elicit S-type CPEs share sequence identity with TcsL-GTD in these same helices. Previous work has shown that mutations—including E449K—alter GTPase profile. Overall, the sequences from all *C. diffiicle* strains indicate a higher rate of amino acid variation in TcdB compared to TcdA.

Currently we are unsure how these rates of mutation compare with the genome average. Analyzing these differences may indicate selection events on either of the toxins. However, significantly more sequences and knowledge of the chromosomal average mutation rate are needed for a complete analysis.

Given that these toxins are so large and thereby place a significant metabolic burden on the cell, many have questioned why *C. difficle* (and *C. sordellii*) maintain both toxins. Regarding $TcdA^-/B^+$ *C. difficile* strains, we do not understand what selective pressures determine the retention or loss of *tcdA*. One hypothesis is that loss of TcdA may improve *C. difficile* competitiveness by simply freeing up the amino acids and energy consumed in translating such a massive protein. An alternate hypothesis is that the inflammatory and antigentic properties of TcdA reduces *C. difficile* fitness during human infection. If the gene is lost, it is less likely for the host to have strong innate and adaptive immune responses. As described in Chapter 2, TcdA has significantly less variation across species compared to TcdB. Low titers of antibodies targeting TcdA and TcdB have been associated with higher recurrence rates and worse patient outcomes[89,90]. Losing this antigenically consistent protein may allow $TcdA^-/TcdB^+$ *C. difficile* strains to avoid elimination across hosts. Finally, recent work done by Lyras et al. has shown that $TcdA^-/TcdB^+$ infection in animal models are significantly more virulent than strains producing both toxins[33]. In part, this work suggests that TcdA may somehow modulate the toxicity of TcdB during infection.

In the course of trying to get co-crystals with this inhibitor, we determined five different structures of the TcdA and TcdB glucosyltransferase domains and made use of a non-hydrolyzable UDP-glucose substrate. This study was initiated as part of an effort

to identify the mode of inhibition for a small molecule inhibitor of glucosyltransferase activity called apigenin. In Chapter 3, I detailed the use of a UPG analogue in crystallizing TcdA and TcdB-GTDs. The use of U2F was essential in obtaining a crystal structure of TcdB-GTD in complex with a non-hydrolyzed substrate. U2F was insufficient to drive conformational changes based on the altered chemical properties of the deoxyfluoroglucose.

The inclusion of apigenin was required for crystallizing an *apo*-like form of TcdB-GTD. However, apigenin was not observed in other TcdA and TcdB-GTD crystals, despite a variety of approaches. In my dataset of the GTD + apigenin crystals, the bridged helices in the TcdB-GTD structure were closely scrutinized in reviewing all TcdA-GTD datasets. In TcdB-GTD, the G444 residues from each ASU chain are located close apigenin. TcdA-GTD, however, has S443 in the same position, which may explain why apigenin is absent in these structures. I suggest apigenin functions to bridge TcdB-GTD monomers as observed in the *apo* ASU. By forming these aggregates, the active site and putative GTPase binding site would be partially occluded, significantly reducing glucosyltransfer rates. Data from Tam et al. clearly show noncompetitive inhibition a by phloretin of the glucosyltransferase activity TcdA and TcdB-GTD[39]. Personal correspondence led us to focus on apigenin, which has approximately five-fold lower K_i compared to phloretin. Previous work has found that apigenin interacts or modulates cystic fibrosis transmembrane conductance regulator (CFTR), IGFBP-3, iNOS, COX-2, glucose transporters, tubulin and other proteins with half-maximal effects (i.e., K_i, K_D, EC/IC_{50}) typically in the 1-100 μM range[91–96]. Additionally, apigenin and phloretin can function as antioxidants, which could improve cellular resilience to the necrotic

mechanism of TcdB[97–100]. Based on its molecular promiscuity it is unlikely that the protection afforded by apigenin is specific to glucosyltransferase activity.

Overall, my work contributes to the study of LCT-GTDs by identifying promising methods for crosslinking and providing new structures of TcdA and TcdB-GTD using a non-hydrolyzable analog. In determining the five new structures of TcdA and TcdB-GTD we observed disparate conformations of the active site tryptophan. These conformations add to the existing collection of LCT-GTD crystals structures where this tryptophan samples a broad conformational space from *apo* to substrate-bound conditions (Figure 4-1).

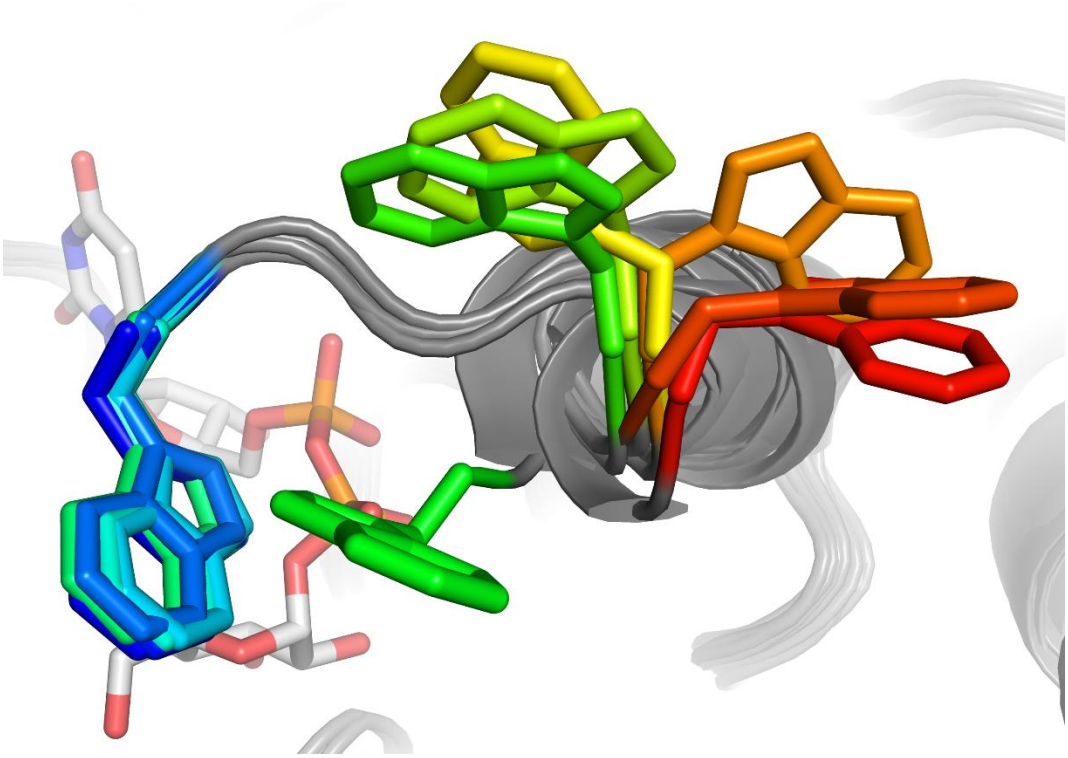

Figure 4-1. Collage of tryptophan conformations observed in crystal structures of LCT-GTDs.

4-2 Future Directions

4-2-1 Evaluate the protective effect of gluconolactone

Due to the abundance and importance of sugars in biology, various natural and synthetic compounds have been tested for glycosyltransferase inhibition or alteration. As discussed in Chapter 2, a glycosyltransferase inhibitor was used to obtain co-crystal structure of TcsL-GTD and UDP[101]. Gluconolactone is another compound with demonstrated inhibition of TcsL-GTD[102]. The lactone form of gluconic acid, gluconolactone has a significantly more electrophilic C1 compared to glucose. Given the hypothetical mechanism of the GTDs (Chapter 1-5, Glycosyltransfer), this enhanced electrophilic nature should more closely approximate the transition state or the UDP-activated sugar. Close to neutral pH, gluconolactone has a relatively short half-life of 10 min[103], which makes co-crystallization unfeasible. Several co-crystal structures containing gluconolactone in the PDB were obtained after soaking crystals in a cryoprotectant solution containing ~30% w/v gluconolactone. Interestingly, gluconolactone appeared to perform quite well as cryoprotectant by itself.

Following these methods, soaked several crystals using a cryoprotectant solution of 50 mg gluconolactone mixed with 100 uL mother liquor. This mixture was aggressively mixed and crushed using a pipette tip to encourage dissolution of crystalline gluconolactone. This was insufficient to dissolve the entire 50 mg. The supernatant of this saturated solution was used as the cryoprotectant and TcdB-GTD + apigenin crystals were soaked for up to six minutes. Diffraction data from these crystals were of average quality for the TcdB-GTD + apigenin crystal form: 3-3-4 Å. However, even in these medium resolution datasets, prominent electron density appears in the

glucose-binding site in both chains within the ASU. The difference map indicated a peak intensity of 11 σ, with a volume and geometry consistent with gluconolactone. Therefore, I have determined that gluconolactone can function as a cryoprotectant and be soaked into at least one crystal form of TcdB-GTD. Acquiring higher resolution data (by soaking UDP-bound TcdB-GTD or *apo* TcdA-GTD crystals) may offer new insights as to the conformation of important active site residues and conformation of gluconolactone. It would also be interesting to examine any reaction or change to gluconolactone within the active site over time. Several intermediate states of gluconolactone and enzyme interactions have been described by ^{1}H-NMR and hydrogen-deuterium exchange[104–106]. These have distinct geometries, which may be distinguishable in a high-resolution crystal structure.

4-2-2 Determine the crystal structure of an APD covalently linked to ebselen

Of the small molecules shown to protect cells from TcdA/TcdB, the drug ebselen has been shown to have effects on both the GTD and APD[42]. Ebselen has pharmacological uses in several psychological disorders including bi-polar disorder. It has also attracted significant interest as a neuroprotective or even antimicrobial drug. Mass spectrometry of ebselen-treated toxin or APD, indicated that the active site cysteine (TcdA C698; TcdB C700) covalently bonds to the organoselenium moiety. Ebselen's nonspecific reactivity to thiols is well-documented. This bicyclic compound covalently bound in the APD active site disrupts normal autoprocessing of the GTD and leads to a delayed CPE in treated cells. This delay is also observed in point mutations of APD cysteines and proteolysis site mutations in the GTD (e.g., TcdB L543A). This autoprocessing defect reduces the ability of the GTD to glucosylate its targets within the

cell. It has also been demonstrated that autoprocessing mutants of TcsL have changes in GTPase glycosylation. With deficient autoprocessing, the GTD would spend a significantly longer period of time anchored to the endosomal membrane of the cytosol, which limits its contact with GTPases. Given enough time, a substantial amount of GTD is released or sufficient GTPases are modified to elicit the CPE. Obtaining a crystal structure of either TcdA- or TcdB-APD would advance our understanding of the active sites. More importantly, the information gained from a crystal structure would aid rational drug design to improve the specificity and efficacy of ebselen as a potential drug against TcdA and TcdB.

4-2-3 Determine the co-crystal structure of GTD and ebselen

When cells were treated with an autoprocessing deficient TcdB (L543A) and ebselen, the CPE was nearly eliminated[42]. This suggests that ebselen has additional effects on LCT intoxication, independent of modifying the APD. A second research team independently identified ebselen as an inhibitor of the toxin, but they indicated it acted by inhibiting the GTD[107]. The inhibition persisted in a cysteine-free construct, which indicates that ebselen's inhibition of the GTDs is independent of thiol reactivity. In order to establish the binding site and mechanism of inhibition, I propose co-crystallizing TcdA or TcdB with ebselen. Given that ebselen is sparingly soluble even in DMSO, obtaining a saturated solution may prove difficult.

4-2-4 Co-crystallize a GTD-GTPase complex

Prior to and during my thesis work, several attempts were made at co-crystallizing TcdA-GTD with RhoA and Rac1. These screens were performed at a molar ratio of 1:1, but failed to yield conditions leading to a co-crystal. Crystals produced from the screens contained TcdA-GTD (most frequently) or a GTPase. Additional screens performed using ligands and substrates including UDP-glucose, UDP, glucose, GTP, GDP, $MnCl_2$, and $MgCl_2$ produced crystals of either protein, but not a co-crystal of the complex. Several alternate approaches may prove more fruitful in obtaining crystals of the complex.

Given the success in crystallizing TcdB-GTD with non-hydrolyzed U2F, this compound should be evaluated in co-crystallization trials of either GTD with GTPases. Since the compound is not hydrolyzed in the absence of substrate, it may promote a more stable complex of GTD and GTPase. However, the U2F compound was difficult for experienced chemists to synthesize which is reflected in its high cost. Advances in enzyme-assisted NDP-sugar synthesis methods could significantly cut costs and complexity[108,109]. However, instead of relying on exotic compounds, a far more tractable approach is site-directed mutagenesis.

Using analogous means of slowing or eliminating the enzymatic turnover with mutants in the GTDs (i.e., D270N, W520A) or GTPases (T35A) may also drive the complex to form using UDP-glucose. In anticipation of these experiments, I have made W519/520A mutations in TcdA and TcdB as well as Switch 1 T35A/T37A mutations in RhoA, Rac1, and Cdc42 expression plasmids. I have successfully expressed and purified the T37A mutant of Rac1.

The problem of diffusion and entropy in this protein-protein complex may be solved *in vivo* by both proteins' mechanisms of membrane localization. *In vitro*, this barrier may be eliminated by constructing a single polypeptide containing both proteins separated by a flexible linker. These fusion proteins have been successfully used in obtaining co-crystal structures of some protein-protein and protein-peptide complexes. I have constructed a plasmid for *B. megaterium* expression of a N-term 6x-His tagged TcdA-GTD-*cum*-Rac1 (T37A) with a 15-residue long linker comprised of GGS repeats. I anticipate this fusion protein will make an excellent candidate for crystallization. In the case this construct does not yield high quality crystals or x-ray data, the Rac1 coding sequence can be truncated to produce Rac1 peptides of any length. This GTD-Rac1 peptide fusion may reduce some instability of a full GTPase fusion and be a better candidate for crystallization. Obtaining high quality data from this construct would yield information about the interactions between the GTPase Switch 1 region and the cognate interface on the GTD. While it would provide less information than a co-crystal structure with a full-length GTPase, the structural data would remain significant[110,111].

4-2-5 Establish the role of the MLD in GTD kinetics

All of the LCT-GTDs contain an N-terminal helical bundle, which has been demonstrated to drive GTD localization to membranes. In some cases, these MLDs preferentially bind to lipid head groups[26,112]. Similarly, most GTPases are lipidated at their C-termini, which allows them to cycle between membrane and cytosolic compartments[113]. It is thought that this co-localization on the lipid bilayer significantly enhances the rate of glucosylation. I have previously attempted to obtain a K_D for the TcdB-GTD and Rac1 complex using microscale thermophoresis (MST). Initial data

suggest the interaction is quite weak, with a K_D greater than 10 μM. The data quality can likely be improved by optimizing the MST buffer conditions (e.g., > 0.2% Tween-20). These data were collected using full-length TcdB-GTD (1-543) and truncated Rac1 (1-181) constructs. The presence of the MLD on the GTD may significantly alter the binding affinity at equilibrium. It has been previously shown that the GTD's ability to localize to membranes can significantly alter the enzyme's access to different GTPases[19]. It may prove useful to generate GTD constructs with MLD deletions or mutations which impair lipid binding. Alternatively, kinetic studies could be performed with lipidated GTPase and full length GTDs in a membrane-associated environment (e.g., using liposomes or giant unilamellar vesicles). Studying kinetics on a lipid bilayer would eschew the need for MLD deletion; it has been shown that one or two point mutations are sufficient to eliminate membrane binding[19].

Ultimately, this work contributes to the development of biological and/or synthetic molecules targeted against the LCT family glycosyltransferases. The topological variation and similarities identified between TcsL-GTD and the variant TcdB-GTDs highlights the importance of the concave face and helices 15, 16, and 17 in particular. In addition, the crosslinking work indicates that a significant amount of a covalent complex can be obtained. A defined interface between these proteins would significantly improve pharmacological efforts to disrupt this interface and reduce LCT lethality. The use of U2F in crystallization efforts demonstrates that UDP-sugar derivatives can approximate ligands and could be adapted to inhibit LCT-GTDs.

LIST OF PUBLICATIONS

Alvin, J. W. & Lacy, D. B. in *Microbial Toxins* (eds. Gopalakrishnakone, P., Stiles, B., Alape-Girón, A., Dubreuil, J. D. & Mandal, M.) 1–18 (Springer Netherlands, 2017).

Alvin, J. W. & Lacy, D. B. Clostridium difficile toxin glucosyltransferase domains in complex with a non-hydrolyzable UDP-glucose analogue. J. Struct. Biol. 198, 203–209 (2017).

BIBLIOGRAPHY

1. Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H. & Kuijper, E. J. *Clostridium difficile* infection. *Nat. Rev. Dis. Prim.* **2,** 16020 (2016).

2. Desai, K. *et al.* Epidemiological and economic burden of *Clostridium difficile* in the United States: estimates from a modeling approach. *BMC Infect. Dis.* **16,** 303 (2016).

3. Alvin, J. W. & Lacy, D. B. in *Microbial Toxins* (eds. Gopalakrishnakone, P., Stiles, B., Alape-Girón, A., Dubreuil, J. D. & Mandal, M.) 1–18 (Springer Netherlands, 2017). doi:10.1007/978-94-007-6725-6_26-2

4. Aronoff, D. M. *Clostridium novyi*, *sordellii*, and *tetani*: Mechanisms of disease. *Anaerobe* **24,** 98–101 (2013).

5. Amimoto, K. *et al.* The protective effect of *Clostridium novyi* type B alpha-toxoid against challenge with spores in guinea pigs. *J. Vet. Med. Sci.* **60,** 681–685 (1998).

6. Coursodon, C. F., Glock, R. D., Moore, K. L., Cooper, K. K. & Songer, J. G. TpeL-producing strains of Clostridium perfringens type A are highly virulent for broiler chicks. *Anaerobe* **18,** 117–121 (2012).

7. Pruitt, R. N. & Lacy, D. B. Toward a structural understanding of *Clostridium difficile* toxins A and B. *Front. Cell. Infect. Microbiol.* **2,** 1–14 (2012).

8. Lowy, I. *et al.* Treatment with monoclonal antibodies against *Clostridium difficile* toxins. *N Engl J Med* **362,** 197–205 (2010).

9. Schorch, B. *et al.* LRP1 is a receptor for *Clostridium perfringens* TpeL toxin indicating a two-receptor model of clostridial glycosylating toxins. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 6431–6 (2014).

10. Yuan, P. *et al.* Chondroitin sulfate proteoglycan 4 functions as the cellular receptor for *Clostridium difficile* toxin B. *Cell Res* **25,** 157–168 (2015).

11. LaFrance, M. E. *et al.* Identification of an epithelial cell receptor responsible for *Clostridium difficile* TcdB-induced cytotoxicity. *Proc. Natl. Acad. Sci.* **112,** 7073–7078 (2015).

12. Chumbler, N. M. *et al.* Crystal structure of *Clostridium difficile* toxin A. *Nat. Microbiol.* **1,** 1–6 (2016).

13. Pruitt, R. N., Chambers, M. G., Ng, K. K.-S., Ohi, M. D. & Lacy, D. B. Structural organization of the functional domains of *Clostridium difficile* toxins A and B. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 13467–72 (2010).

14. Zhang, Z. *et al.* Translocation domain mutations affecting cellular toxicity identify the *Clostridium difficile* toxin B pore. *Proc Natl Acad Sci U S A* **111,** 3721–3726 (2014).

15. Guttenberg, G. *et al.* Inositol hexakisphosphate-dependent processing of *Clostridium sordellii* lethal toxin and *Clostridium novyi* alpha-toxin. *J Biol Chem* **286,** 14779–14786 (2011).

16. Kreimeyer, I. *et al.* Autoproteolytic cleavage mediates cytotoxicity of *Clostridium difficile* toxin A. *Naunyn Schmiedebergs Arch Pharmacol* **383,** 253–262 (2011).

17. Chumbler, N. M. *et al.* Clostridium difficile Toxin B Causes Epithelial Cell Necrosis through an Autoprocessing-Independent Mechanism. *PLoS Pathog.* **8,** e1003072 (2012).

18. Li, S., Shi, L., Yang, Z. & Feng, H. Cytotoxicity of *Clostridium difficile* toxin B does not require cysteine protease-mediated autocleavage and release of the glucosyltransferase domain into the host cell cytosol. *Pathog. Dis.* **67,** 11–18 (2013).

19. Craven, R. & Lacy, D. B. *Clostridium sordellii* Lethal-Toxin Autoprocessing and Membrane Localization Activities Drive GTPase Glucosylation Profiles in Endothelial Cells. *Am. Soc. Microbiol. Msph.* **1,** 1–9 (2015).

20. Genth, H. *et al.* Haemorrhagic toxin and lethal toxin from *Clostridium sordellii* strain vpi9048: Molecular characterization and comparative analysis of substrate specificity of the large clostridial glucosylating toxins. *Cell. Microbiol.* **16,** 1706–1721 (2014).

21. Nagahama, M. *et al. Clostridium perfringens* TpeL glycosylates the Rac and Ras subfamily proteins. *Infect. Immun.* **79,** 905–910 (2011).

22. Ziegler, M. O. P., Jank, T., Aktories, K. & Schulz, G. E. Conformational Changes and Reaction of Clostridial Glycosylating Toxins. *J. Mol. Biol.* **377,** 1346–1356 (2008).

23. Laughlin, M. R., Petit, W. A., Dizon, J. M., Shulman, R. G. & Barrett, E. J. NMR measurements of *in vivo* myocardial glycogen metabolism. *J. Biol. Chem.* **263,** 2285–2291 (1988).

24. Jank, T. & Aktories, K. Structure and mode of action of clostridial glucosylating toxins: the ABCD model. *Trends Microbiol.* **16,** 222–229 (2008).

25. Jank, T., Giesemann, T. & Aktories, K. Rho-glucosylating *Clostridium difficile* toxins A and B : new insights into structure and function. *Glycobiology* **17,** 15–22 (2007).

26. Varela Chavez, C. *et al.* The Tip of the Four N-Terminal α-Helices of *Clostridium sordellii* Lethal Toxin Contains the Interaction Site with Membrane Phosphatidylserine Facilitating Small GTPases Glucosylation. *Toxins (Basel).* **8,** (2016).

27. Halabi-Cabezon, I. *et al.* Prevention of the cytopathic effect induced by *Clostridium difficile* Toxin B by active Rac1. *FEBS Lett.* **582,** 3751–3756 (2008).

28. Rupnik, M. An Update on *Clostridium difficile* Toxinotyping. **54,** 13–18 (2016).

29. Schulz, F., Just, I. & Genth, H. Prevention of *Clostridium sordellii* lethal toxin-induced apoptotic cell death by tauroursodeoxycholic acid. *Biochemistry* **48,** 9002–9010 (2009).

30. Guttenberg, G. *et al.* Molecular characteristics of *Clostridium perfringens* TpeL toxin and consequences of mono-O-GlcNAcylation of Ras in living cells. *J. Biol. Chem.* **287,** 24929–24940 (2012).

31. Farrow, M. A. *et al. Clostridium difficile* toxin B-induced necrosis is mediated by the host epithelial cell NADPH oxidase complex. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 18674–9 (2013).

32. Frädrich, C., Beer, L. A. & Gerhard, R. Reactive oxygen species as additional determinants for cytotoxicity of *Clostridium difficile* toxins A and B. *Toxins (Basel).* **8,** 1–12 (2016).

33. Carter, G. P. *et al.* Defining the roles of TcdA and TcdB in localized gastrointestinal disease, systemic organ damage, and the host response during *Clostridium difficile* infections. *MBio* **6,** 1–10 (2015).

34. Amimoto, K., Noro, T., Oishi, E. & Shimizu, M. A novel toxin homologous to large clostridial cytotoxins found in culture supernatant of *Clostridium perfringens* type C. *Microbiology* **153,** 1198–1206 (2007).

35. Sougioultzis, S. *et al. Clostridium difficile* toxoid vaccine in recurrent *C. difficile*-associated diarrhea. *Gastroenterology* **128,** 764–770 (2005).

36. Beilhartz, G. L., Tam, J. & Melnyk, R. A. Small Molecules Take A Big Step Against *Clostridium difficile* . *Trends Microbiol.* **23,** 746–748 (2015).

37. Tao, L. *et al.* Frizzled proteins are colonic epithelial receptors for *C. difficile* toxin B. *Nature* **538,** 350–355 (2016).

38. Slater, L. H. *et al.* Identification of Novel Host-Targeted Compounds That Protect from Anthrax Lethal Toxin-Induced Cell Death. *ACS Chem. Biol.* **8,** 812–822 (2013).

39. Tam, J. *et al.* Small molecule inhibitors of *Clostridium difficile* toxin B-induced cellular damage. *Chem. Biol.* **22,** 175–85 (2015).

40. Darkoh, C., Brown, E. L., Kaplan, H. B. & DuPont, H. L. Bile salt inhibition of host cell damage by *Clostridium difficile* toxins. *PLoS One* **8,** 1–9 (2013).

41. Shen, A. *et al.* Defining an allosteric circuit in the cysteine protease domain of *Clostridium difficile* toxins. *Nat. Struct. Mol. Biol.* **18,** 364–371 (2011).

42. Bender, K. O. *et al.* A small-molecule antivirulence agent for treating *Clostridium difficile* infection. *Sci. Transl. Med.* **7,** 306ra148 (2015).

43. Smith, S. M. E. *et al.* Ebselen and congeners inhibit NADPH oxidase 2-dependent superoxide generation by interrupting the binding of regulatory subunits. *Chem. Biol.* **19,** 752–763 (2012).

44. Jank, T., Belyi, Y. & Aktories, K. Bacterial glycosyltransferase toxins. *Cell. Microbiol.* **17,** 1752–1765 (2015).

45. Fluit, A. C. *et al.* Nontoxigenic strains of *Clostridium difficile* lack the genes for both toxin A and toxin B. *J. Clin. Microbiol.* **29,** 2666–7 (1991).

46. Zeiser, J., Gerhard, R., Just, I. & Pich, A. Substrate specificity of clostridial glucosylating toxins and their function on colonocytes analyzed by proteomics techniques. *J. Proteome Res.* **12,** 1604–1618 (2013).

47. May, M., Wang, T., Müller, M. & Genth, H. Difference in F-actin depolymerization induced by toxin B from the *Clostridium difficile* strain VPI 10463 and toxin B from the variant *Clostridium difficile* serotype F strain 1470. *Toxins (Basel).* **5,** 106–119 (2013).

48. Chaves-Olarte, E. *et al.* A novel cytotoxin from *Clostridium difficile* serogroup F is a functional hybrid between two other large clostridial cytotoxins. *J. Biol. Chem.* **274,** 11046–11052 (1999).

49. Li, W. *et al.* The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* **43,** W580–W584 (2015).

50. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42,** W320–W324 (2014).

51. Biasini, M. *et al.* SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42,** (2014).

52. Jank, T., Giesemann, T. & Aktories, K. *Clostridium difficile* glucosyltransferase toxin B-essential amino acids for substrate binding. *J. Biol. Chem.* **282,** 35222–35231 (2007).

53. Pruitt, R. N. *et al.* Structural determinants of *Clostridium difficile* toxin A glucosyltransferase activity. *J. Biol. Chem.* **287,** 8013–8020 (2012).

54. Fabrini, R. *et al.* Monomer−Dimer Equilibrium in Glutathione Transferases: A Critical Re-Examination. *Biochemistry* **48,** 10473–10482 (2009).

55. Ellman, J., Mendel, D., Anthony-Cahill, S., Noren, C. J. & Schultz, P. G. in 301–336 (1991). doi:10.1016/0076-6879(91)02017-4

56. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 11020–4 (2002).

57. Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *J. Mol. Biol.* **395,** 361–374 (2010).

58. Alley, S. C., Ishmael, F. T., Jones, A. D. & Benkovic, S. J. Mapping protein - Protein interactions in the bacteriophage T4 DNA polymerase holoenzyme using a novel trifunctional photo-cross-linking and affinity reagent [11]. *Journal of the American Chemical Society* **122,** 6126–6127 (2000).

59. Trakselis, M. A., Alley, S. C. & Ishmael, F. T. Identification and mapping of protein-protein interactions by a combination of cross-linking, cleavage, and proteomics. *Bioconjugate Chemistry* **16,** 741–750 (2005).

60. Alvin, J. W. & Lacy, D. B. Clostridium difficile toxin glucosyltransferase domains in complex with a non-hydrolyzable UDP-glucose analogue. *J. Struct. Biol.* **198,** 203–209 (2017).

61. Hofmann, F., Herrmann, A., Habermann, E. & Von Eichel-Streiber, C. Sequencing and analysis of the gene encoding the α-toxin of *Clostridium novyi* proves its homology to toxins A and B of *Clostridium difficile* . *Mol. Gen. Genet.* **247,** 670–679 (1995).

62. Carter, G. P., Rood, J. I. & Lyras, D. The role of toxin A and toxin B in the virulence of *Clostridium difficile* . *Trends in Microbiology* **20,** 21–29 (2012).

63. Reinert, D. J., Jank, T., Aktories, K. & Schulz, G. E. Structural basis for the function of *Clostridium*

difficile toxin B. *J. Mol. Biol.* **351,** 973–981 (2005).

64. Ciesla, W. P. & Bobak, D. A. *Clostridium difficile* Toxins A and B Are Cation-dependent UDP-glucose Hydrolases with Differing Catalytic Activities. *J. Biol. Chem.* **273,** 16021–16026 (1998).

65. Chaves-Olarte, E., Weidmann, M., Von Eichel-Streiber, C. & Thelestam, M. Toxins A and B from *Clostridium difficile* differ with respect to enzymatic potencies, cellular substrate specificities, and surface binding to cultured cells. *J. Clin. Invest.* **100,** 1734–1741 (1997).

66. Barth, H. *et al.* Low pH-induced Formation of Ion Channels by *Clostridium difficile* Toxin B in Target Cells. *J. Biol. Chem.* **276,** 10670–10676 (2001).

67. Reineke, J. *et al.* Autocatalytic cleavage of *Clostridium difficile* toxin B. *Nature* **446,** 415–419 (2007).

68. Just, I. *et al.* Glucosylation of Rho proteins by *Clostridium difficile* toxin B. *Nature* **375,** 500–503 (1995).

69. Busch, C. *et al.* A common motif of eukaryotic glycosyltransferases is essential for the enzyme activity of large clostridial cytotoxins. *J. Biol. Chem.* **273,** 19566–19572 (1998).

70. Nagahama, M., Oda, M. & Kobayashi, K. *Glycosylating Toxin of* Clostridium perfringens. *Glycosylation* (2012). doi:10.5772/48112

71. D'Orzo, N. *et al.* The structure of *Clostridium difficile* toxin A glucosyltransferase domain bound to Mn 2+ and UDP provides insights into glucosyltransferase activity and product release. *FEBS J.* **279,** 3085–3097 (2012).

72. Gibson, R. P., Tarling, C. A., Roberts, S., Withers, S. G. & Davies, G. J. The donor subsite of trehalose-6-phosphate synthase: Binary complexes with udp-glucose and udp-2-deoxy-2-fluoro-glucose at 2 Å resolution. *J. Biol. Chem.* **279,** 1950–1955 (2004).

73. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 125–132 (2010).

74. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology* **276,** 307–326 (1997).

75. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).

76. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

77. Schrödinger, L. *The PyMOL Molecular Graphics System, Version 1.8*. (2015).

78. Persson, K. *et al.* Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nat. Struct. Biol.* **8,** 166–175 (2001).

79. Truman, A. W. *et al.* Chimeric Glycosyltransferases for the Generation of Hybrid Glycopeptides. *Chem. Biol.* **16,** 676–685 (2009).

80. Hiromoto, T. *et al.* Crystal structure of UDP-glucose:anthocyanidin 3-O-glucosyltransferase from *Clitoria ternatea*. *J. Synchrotron Radiat.* **20,** 894–898 (2013).

81. Busch, C., Hofmann, F., Gerhard, R. & Aktories, K. Involvement of a conserved tryptophan residue in the UDP-glucose binding of large clostridial cytotoxin glycosyltransferases. *J. Biol. Chem.* **275,** 13228–13234 (2000).

82. Spyres, L. M. *et al.* Mutational analysis of the enzymatic domain of *Clostridium difficile* toxin B reveals novel inhibitors of the wild-type toxin. *Infect. Immun.* **71,** 3294–3301 (2003).

83. Teichert, M., Tatge, H., Schoentaube, J., Just, I. & Gerhard, R. Application of mutated *Clostridium difficile* toxin A for determination of glucosyltransferase-dependent effects. *Infect. Immun.* **74,** 6006–6010 (2006).

84. Breton, C., Šnajdrová, L., Jeanneau, C., Koča, J. & Imberty, A. Structures and mechanisms of glycosyltransferases. *Glycobiology* **16,** 29–37 (2006).

85. Liebschner, D. *et al.* Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. Sect. D Struct. Biol.* **73,** 148–157 (2017).

86. Guttenberg, G. *et al.* Inositol hexakisphosphate-dependent processing of Clostridium sordellii lethal toxin and *Clostridium novyi* α-toxin. *J. Biol. Chem.* **286,** 14779–14786 (2011).

87. Kim, M.-S. *et al.* A draft map of the human proteome. *Nature* **509,** 575–581 (2014).

88. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317–330 (2015).

89. Wullt, M., Norén, T., Ljungh, A. & Åkerlund, T. IgG antibody response to toxins A and B in patients with *Clostridium difficile* infection. *Clin. Vaccine Immunol.* **19,** 1552–4 (2012).

90. Kyne, L. *et al.* Association between antibody response to toxin A and protection against recurrent Clostridium difficile diarrhoea. *Lancet (London, England)* **357,** 189–93 (2001).

91. Shukla, S. Up-regulation of insulin-like growth factor binding protein-3 by apigenin leads to growth inhibition and apoptosis of 22Rv1 xenograft in athymic nude mice. *FASEB J.* **19,** 2042–4 (2005).

92. Liang, Y.-C. *et al.* Suppression of inducible cyclooxygenase and inducible nitric oxide synthase by apigenin and related flavonoids in mouse macrophages. *Carcinogenesis* **20,** 1945–1952 (1999).

93. Illek, B. *et al.* Structural determinants for activation and block of CFTR-mediated chloride currents by apigenin. *Am. J. Physiol. - Cell Physiol.* **279,** (2000).

94. Choudhury, D. *et al.* Apigenin shows synergistic anticancer activity with curcumin by binding at different sites of tubulin. *Biochimie* **95,** 1297–1309 (2013).

95. Melstrom, L. G. *et al.* Apigenin inhibits the GLUT-1 glucose transporter and the phosphoinositide 3-kinase/Akt pathway in human pancreatic cancer cells. *Pancreas* **37,** 426–31 (2008).

96. Ren, B. *et al.* Apigenin and naringenin regulate glucose and lipid metabolism, and ameliorate vascular dysfunction in type 2 diabetic rats. *Eur. J. Pharmacol.* **773,** 13–23 (2016).

97. Han, J.-Y. *et al.* Protection of apigenin against kainate-induced excitotoxicity by anti-oxidative effects. *Biol. Pharm. Bull.* **35,** 1440–6 (2012).

98. Wang, Y. C. & Huang, K. M. In vitro anti-inflammatory effect of apigenin in the Helicobacter pylori-infected gastric adenocarcinoma cells. *Food Chem. Toxicol.* **53,** 376–383 (2013).

99. San Miguel, S. M., Opperman, L. A., Allen, E. P., Zielinski, J. & Svoboda, K. K. H. Bioactive polyphenol antioxidants protect oral fibroblasts from ROS-inducing agents. *Arch. Oral Biol.* **57,** 1657–1667 (2012).

100. Zuo, A. R. *et al.* Hepatoprotective effects and antioxidant, antityrosinase activities of phloretin and phloretin isonicotinyl hydrazone. *J. Chinese Med. Assoc.* **77,** 290–301 (2014).

101. Jank, T., Ziegler, M. O. P., Schulz, G. E. & Aktories, K. Inhibition of the glucosyltransferase activity of clostridial Rho/Ras-glucosylating toxins by castanospermine. *FEBS Lett.* **582,** 2277–2282 (2008).

102. Geyer, M., Wilde, C., Selzer, J., Aktories, K. & Kalbitzer, H. R. Glucosylation of Ras by *Clostridium sordellii* Lethal Toxin: Consequences for Effector Loop Conformations Observed by NMR Spectroscopy. *Biochemistry* **42,** 11951–11959 (2003).

103. Van Dijken, J. P., Van Tuijl, A., Luttik, M. A. H., Middelhoven, W. J. & Pronk, J. T. Novel pathway for alcoholic fermentation of ??-gluconolactone in the yeast Saccharomyces bulderi. *J. Bacteriol.* **184,** 672–678 (2002).

104. Wałaszek, Z., Horton, D. & Ekiel, I. Conformational studies on aldonolactones by n.m.r. spectroscopy. Conformations of d-glucono-1,5-lactone and d-mannono-1,5-lactone in solution. *Carbohydr. Res.* **106,** 193–201 (1982).

105. Sinnott, M. L. Catalytic mechanism of enzymic glycosyl transfer. *Chem. Rev.* **90,** 1171–1202 (1990).

106. Kim, S. C., Singh, A. N. & Raushel, F. M. The mechanism of glycogen synthetase as determined by deuterium isotope effects and positional isotope exchange experiments. *J. Biol. Chem.* **263,** 10151–4 (1988).

107. Beilhartz, G. L., Tam, J., Zhang, Z. & Melnyk, R. A. Comment on 'A small-molecule antivirulence agent for treating *Clostridium difficile* infection'. *Sci. Transl. Med.* **8,** (2016).

108. Shaikh, F. A. & Withers, S. G. Teaching old enzymes new tricks: engineering and evolution of glycosidases and glycosyl transferases for improved glycoside synthesisThis paper is one of a selection of papers published in this Special Issue, entitled CSBMCB — Systems and Chemical Biology. *Biochem. Cell Biol.* **86,** 169–177 (2008).

109. Liu, Z., Zhang, J., Chen, X. & Wang, P. G. Combined biosynthetic pathway for de novo production of UDP-galactose: Catalysis with multiple enzymes immobilized on agarose beads. *ChemBioChem* **3,** 348–355 (2002).

110. Reddy Chichili, V. P., Kumar, V. & Sivaraman, J. Linkers in the structural biology of protein-protein interactions. *Protein Sci.* **22,** 153–167 (2013).

111. Kumar, V. *et al.* Structural Basis for the Interaction of Unstructured Neuron Specific Substrates Neuromodulin and Neurogranin with Calmodulin. *Sci. Rep.* **3,** 1392 (2013).

112. Geissler, B., Tungekar, R. & Satchell, K. J. F. Identification of a conserved membrane localization domain within numerous large bacterial protein toxins. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 5581–5586 (2010).

113. Genth, H., Aktories, K. & Just, I. Monoglucosylation of RhoA at threonine 37 blocks cytosol-membrane cycling. *J. Biol. Chem.* **274,** 29050–29056 (1999).