Clostridium difficile Toxins: Entry and Uptake

by

Zhifen Zhang

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Department of Biochemistry
University of Toronto

© Copyright by Zhifen Zhang 2018
Abstract

*Clostridium difficile* is a major nosocomial pathogen that causes hospital-acquired and antibiotic-associated diarrhea through the actions of two related exotoxins (TcdA and TcdB) on the host. They are large, multi-domain toxins that enter cells by receptor-mediated endocytosis. Upon acidification, embedded hydrophobic segments within the translocation domain insert into the endosomal membrane to form a transmembrane pore, via which toxic enzymatic domain is thought to insert and translocate into the cytosol to target small GTPases within the host. Our understanding of the cell surface receptors and the mechanisms of pore formation and translocation of these toxins has been limited for many years by a lack of any structural information for the translocation domain. This thesis focuses on identification and elucidation of key determinants of pore formation within the central ~1000 amino acid translocation domain and on receptor binding. Through a systematic perturbation of conserved sites within the predicted 172-residue hydrophobic region of the translocation domain, we identified crucial residues for pore-formation and uncovered unexpected similarities to the diphtheria toxin translocation domain, in terms of hydrophobicity patterning and the identity of key pore-forming residues, leading us to propose a “double-dagger” model of membrane insertion. The high-

*Clostridium difficile* Toxins: Entry and Uptake

Zhifen Zhang

Doctor of Philosophy

Department of Biochemistry
University of Toronto

2018
resolution structure for TcdA in the pre-pore conformation further revealed a novel way of packing hydrophobic segments for *C. difficile* toxins, in which the majority of the translocation domain served as the scaffold for the hydrophobic region, which was splayed around it, rather than being buried inside the protein. Additionally, our systematic screen also led to the unexpected identification of a small region at the junction of the translocation domain and the C-terminal combined repetitive oligopeptides (CROP) that was important for function. We discovered that this region was the elusive binding site for one of the TcdB receptors, Chondroitin Sulfate Proteoglycan 4 (CSPG4). Using a series of truncations, we showed that TcdB required only three short CROP repeats for CSPG4 binding, resolving some contradictory data regarding role of CROP domain in the field. Overall, the results of these studies provide mechanistic insights into structure and mechanism of entry and uptake for *C. difficile* toxins.
Acknowledgments

This thesis represented immense support and opportunities I have fortunately encountered over the past few years. I would like to express my sincere gratitude towards all those people who have contributions to the completion of the thesis and supported me through the whole journey.

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Roman Melnyk. I am profoundly fortunate to be the first student and mentored by him over the past six years. Thank you for your guidance, support, and encouragement and for sharing your wisdom both in research and life. Thank you for always listening to my problems, guiding me to the correct path, and at the same time giving me freedom to pursue my ideas and hypothesis. I have truly enjoyed working in your lab and have learned so much from you and grown throughout my graduate years.

I would also like to express my appreciation towards my supervisory committee members, Dr. John Brumell and Dr. Trevor Moraes, for their time, invaluable advices and encourage throughout all my committee meetings. I would also like to thank all my collaborators, especially Dr. Borden Lacy and Dr. Pulkit Gupta from Merck co., which all helped tremendously to the completion of my work. I enjoyed the numerous discussions over the years, which helped me progress in my projects.

All these work would not be possible without the help and support from all the wonderful members from the Melnyk lab, John Tam, Anick Auger, Dr. Greg Beilhartz, Dr. Seiji Sugiman-Marangos, Minyoung Emily Park, Kathleen Orrell and Simoun Icho. I especially thank John and Anick for helping me and guiding me step by step in the beginning from simple cloning to thinking critically and trouble shooting. I have learned a great deal from both of you; the screening would not be possible without your help. Thank you Greg and Seiji for all these interesting discussions and providing constructive suggestions both in life and research whenever I was in doubt or lost. You two are the most intelligent and dedicated scientist I know and your enthusiasm towards science is always an inspiration for me. Many thanks to Minyoung (Emily), my “doppelgänger”, for just being there whenever I was down and listening to my problems and providing the comfort through all my up and downs during graduate school. I would like to thank Katheleen and Simoun for continuing C. difficile projects and your help during
preparations of my thesis. It is my privilege to work with all of you and thank you for making this lab another home for me.

Last but not the least, I would like to extend my deepest thanks to my parents for their support through all my years of study in Canada. Thank you for being there during difficult times and providing me with unconditional love and support.
Table of Contents

Acknowledgments........................................................................................................ iv
Table of Contents........................................................................................................ vi
List of Abbreviations .................................................................................................. xii
List of Figures ................................................................................................................ xv
List of Tables ................................................................................................................ xvii
List of Appendices ....................................................................................................... xviii
Chapter 1 ....................................................................................................................... 1
  1 Introduction ................................................................................................................. 2
    1.1 Clostridium difficile: Historical Perspective ......................................................... 2
    1.2 Clostridium difficile Pathogenicity ....................................................................... 2
      1.2.1 Clostridium difficile associated diseases ...................................................... 2
      1.2.2 Clostridium difficile lifecycle ........................................................................ 3
      1.2.3 Regulation of toxin production ..................................................................... 4
    1.3 Clostridium difficile toxins ..................................................................................... 5
      1.3.1 Role of TcdA and TcdB in CDI ................................................................. 5
      1.3.2 Structure and function of TcdA and TcdB ................................................... 6
        1.3.2.1 Cellular binding and uptake ................................................................ 10
          1.3.2.1.1 The CROP domain ....................................................................... 10
          1.3.2.1.2 Role of CROP domain in C. difficile toxins ............................... 12
          1.3.2.1.3 TcdA receptors ............................................................................. 12
          1.3.2.1.4 TcdB receptors ............................................................................. 13
          1.3.2.1.5 CROP-targeted neutralizing antibodies prevent receptor-binding .... 13
        1.3.2.2 Endocytosis ............................................................................................. 15
        1.3.2.3 Pore formation and translocation ......................................................... 16
1.3.2.3.1 TcdA and TcdB undergo pH-dependent conformational changes that are required for toxicity ........................................16
1.3.2.3.2 Structural determinants for pore formation ........................................18
1.3.2.4 Autoprocessing (APD) ..............................................................................19
1.3.2.5 Biological activity of C. difficile toxins .................................................21
  1.3.2.5.1 Structure and function of glucosyltransferase domain (GTD) ......21
  1.3.2.5.2 Rho modification by GTD .................................................................23
1.3.3 Cellular effects of TcdA and TcdB ...............................................................25
  1.3.3.1 Cytopathic effects ..................................................................................25
  1.3.3.2 Cytotoxic/Apoptosis Effects ..................................................................26
  1.3.3.3 Glucosylation-independent necrosis effects ........................................26
  1.3.3.4 Toxin induced inflammation and its role in diseases .......................27
1.4 Existing Problems and Thesis Rational ............................................................28

Chapter 2 ...............................................................................................................31

2 Translocation Domain Mutations Affecting Cellular Toxicity Identify the Clostridium difficile Toxin B Pore ...........................................................................32
  2.1 Overview ........................................................................................................32
  2.2 Introduction ....................................................................................................32
  2.3 Material and Methods ..................................................................................33
    2.3.1 Production of TcdB ...................................................................................33
      2.3.1.1 Expression and purification of recombinant TcdB from Bacillus megaterium .............................................................................................33
      2.3.1.2 Expression and purification of recombinant codon-optimized TcdB constructs .........................................................................................34
      2.3.1.3 Mutagenesis of TcdB mutants ...................................................................34
      2.3.1.4 Small-scale expression of TcdB mutants ................................................34
    2.3.2 Characterization Assays ...........................................................................34
      2.3.2.1 Cell viability assay ................................................................................34
2.3.2.2 CellTiterGlo ATP assay. ................................................................. 35
2.3.2.3 Rubidium release assay. ............................................................... 35
2.3.2.4 TNS fluorescence assay ............................................................... 36
2.3.2.5 Black lipid bilayer experiments .................................................. 36

2.4 Results ................................................................................................. 36

2.4.1 Patterns of Hydrophobicity and Secondary Structure Suggest a Helical Pore for TcdB ................................................................. 36
2.4.2 Validation of a GC-Enriched Toxin B Gene for Mutagenesis and Expression in E. coli ................................................................. 41
2.4.3 High-Throughput Mapping of the Functional Determinants in the Translocation Domain................................................................. 41
2.4.4 Pinpointing the Nature of the Defects in TcdB that Diminish Function ........ 43
2.4.5 Examining the Role of Pore-Formation on TcdB-Induced Necrosis .......... 45
2.4.6 Mapping the Determinants of Pore formation onto a Model of the TcdB Pore ... 46

2.5 Discussion ............................................................................................ 47

Chapter 3 ................................................................................................. 50

3 Defective Mutations within the Translocation Domain of Clostridium difficile Toxin B Impair Disease ................................................................. 51

3.1 Overview ............................................................................................ 51
3.2 Introduction ........................................................................................ 51

3.3 Material and methods ......................................................................... 51

3.3.1 Mouse systemic toxin challenge .................................................... 51
3.3.2 Differential scanning fluorometry .................................................. 52
3.3.3 Glucosyltransferase activity of the toxins ....................................... 52

3.4 Results ................................................................................................. 53

3.4.1 Generation of the functional residues at amino acid 1106 ................. 53
3.4.2 L1106K mutation diminishes its activity in vivo ......................... 54
3.4.3 Characterization of L1106X activities in vitro ............................... 54
3.5 Discussion.................................................................................................................. 56

Chapter 4.......................................................................................................................... 58

4 Crystal Structure of *Clostridium difficile* Toxin A .................................................... 59

4.1 Overview...................................................................................................................... 59

4.2 Material and Methods .............................................................................................. 59

4.2.1 Protein expression and purification ................................................................. 59

4.2.2 Crystallization ...................................................................................................... 60

4.2.3 Structure determination and refinement .......................................................... 60

4.2.4 X-ray absorption spectroscopy .......................................................................... 61

4.2.5 ICP-MS ................................................................................................................ 62

4.2.6 Autoprocessing assays ....................................................................................... 62

4.2.7 Cell lines .............................................................................................................. 63

4.2.8 Cell binding and Rac1 glucosylation in cells...................................................... 63

4.2.9 Statistical analyses .............................................................................................. 63

4.3 Results and Discussion ............................................................................................ 63

Chapter 5.......................................................................................................................... 72

5 Functional Defects in *Clostridium difficile* TcdB Toxin Uptake Identify CSPG4 Receptor Binding Determinants .............................................................. 73

5.1 Overview .................................................................................................................... 73

5.2 Introduction ............................................................................................................... 73

5.3 Materials and Methods ............................................................................................ 75

5.3.1 Production of TcdB proteins .............................................................................. 75

5.3.1.1 Cloning and mutagenesis ............................................................................. 75

5.3.2 Cell Culture and generation of cell lines ............................................................ 75

5.3.2.1 Cell lines and reagents ................................................................................ 75

5.3.2.2 Generation of CSPG4 expressing HEK 293 Cell Line ................................ 76
5.3.2.3 CRISPR mutagenesis.................................................................76
5.3.3 Characterization of protein constructs ........................................77
  5.3.3.1 Toxin cell death assay ..........................................................77
5.3.4 Binding Assays ...........................................................................77
  5.3.4.1 TcdB cell surface binding assay .............................................77
  5.3.4.2 Co-immunoprecipitation assays .............................................78
  5.3.4.3 Immunoblotting ....................................................................78
  5.3.4.4 Fzd7 purification .................................................................79
  5.3.4.5 Gel filtration assay ...............................................................79
5.4 Results ..........................................................................................79
  5.4.1 Mutations at the boundary of the translocation and the CROPs domain affect TcdB function ..........................................................79
  5.4.2 Defective mutants show decreased host cell-surface binding ..........81
  5.4.3 Defective TcdB mutants decrease binding to CSPG4, but not NECTIN3 or FZD7 .................................................................83
  5.4.4 TcdB requires three oligopeptide repeats for CSPG4-binding and full cellular activity .............................................................85
  5.4.5 The CROP domain from TcdB is not sufficient for cell surface binding ....87
  5.4.6 Antibody binding to TcdB CROPs occludes CSPG4 binding ..........89
5.5 Discussion .....................................................................................92
Chapter 6 ............................................................................................95
6 Conclusion, Discussion and Future Directions .......................................96
  6.1 Mechanism of Pore-formation and Translocation .............................96
    6.1.1 The link between C. difficile toxins and other pore-forming toxins ....98
    6.1.2 Requirements for TcdA and TcdB translocation machinery ..........101
    6.1.3 Pore formation vs. Translocation .............................................103
    6.1.4 Future directions .....................................................................105
6.2 Receptor binding ........................................................................................................................................107
  6.2.1 Role of CROP domain in TcdA and TcdB .......................................................................................107
  6.2.2 What is the physiological relevance of the multiple receptor model? .............................................108
  6.2.3 Future directions: elucidating TcdB receptor binding sites ..............................................................109
  6.2.4 Identifying Receptor binding for TcdA ............................................................................................110
6.3 The BioID method to elucidate binding and translocation ......................................................................111
6.4 Concluding Remarks ................................................................................................................................111
Appendices .........................................................................................................................................................113
References .........................................................................................................................................................125
List of Abbreviations

APD   Autoprotease domain
ATPase Adenosine triphosphatase
B.meg Bacillus megaterium
BAX   BCL-2-like protein 4
BCL-2 B-cell lymphoma 2
BioID Proximity-dependent biotin identification
C. difficile Clostridium difficile
Cas9  CRISPR associated protein 9
CDAD Clostridium difficile associated disease
CDAI Clostridium difficile infections
CHO-K1 Chinese hamster ovary cell
CPD   Cysteine protease domain
Cq    Quantification cycle
CRD   Cysteine-rich domain
CRISPR Clustered regularly interspaced short palindromic repeats
CROP Combined repetitive oligopeptide
CspC  Clostridium difficile bile acid germinant receptor
CSPG4 Chondroitin sulfate proteoglycan 4
Cwp84 Cell wall protein 84
Cys   Cysteine
DMEM  Dulbecco’s minimal essential medium
DT    Diphtheria toxin
E.coli Escherichia coli
EC50  Half-maximum effective concentration
ECF   Extracytoplasmic Function
EM    Electron microscopy
EMEM  Eagle’s Minimal essential medium
FBS   Fetal calf serum
FZD   Wnt receptor frizzled protein family
GAPs  GTPase activating proteins
GDIs  Guanine dissociation inhibitors
GDP   Guanosine diphosphate
GEFs  Guanine nucleotide exchange proteins
GlcNAc N-Acetylg glucosamine
GT-A  Glucosyltransferase type A
GTD   Glucosyltransferase domain
GTP   Guanosine triphosphate
GTPase Guanosinetriphosphatase
HEPES N-2-Hydroxyethylpiperazine-N’-2-Ethanesulfonic Acid
HH    Hydrophobic helices
His   Histidinie
HIV   Human immunodeficiency virus
HRP   Horseradish peroxidase
ICP-MS Inductively coupled plasma-mass spectrometry
IL-6/8 Interleukin-6/8
InsP6  inositol hexakisphosphate
IP     Intraperitoneally
IPTG  Isopropyl β-D-1-thiogalactopyranoside
ISEMF Intestinal sub-epithelial myofibroblast
LB    Luria-Bertani
LCTs  Large clostridial toxins
LD    Lethal dose
LDH   Lactate dehydrogenase
LR    Long repeat
Lys   Lysine
MARTX Multifunctional-autoprocessing-repeats-in-toxin
MLD   Membrane localization domain
Mut   Mutant
MW    Molecular weight
NADPH Nicotinamide adenine dinucleotide phosphate
NAP1/027 North American pulsed-field gel electrophoresis type 1
NOX   NADPH oxidase
PACSIN2 Protein kinase C and casein kinase substrate in neurons 2
PaLoc Pathogenicity locus
PCR   Polymerase chain reaction
PDB   Protein data bank
PLB   Planar lipid bilayer
PMC   Pseudomembranous colitis
PVRL3/NECTIN3 Poliovirus receptor-like 3
qPCR  Quantitative real-time polymerase chain reaction
Rb    Rubidium
RBD   Receptor binding domain
RM    Rough microsome
RNA   Ribonucleic acid
RNAi  Ribonucleic acid interference
ROS   Reactive oxygen species
SDS-PAGE Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SR    Short repeat
SRB   Sulforhodamine B
TcdA  Clostridium difficile toxin A
TcdB  Clostridium difficile toxin B
TCEP Tris(2-carboxyethyl)phosphine
TcnA  Clostridium novyi alpha-toxin
TcsH  Clostridium hemorrhagic toxin
TcsL  Clostridium sordellii lethal toxin
TD    Translocation domain
TH    Transmembrane helices
TM    Transmembrane
TMHMM Membrane protein topology prediction method
TNS  2-(p-toluidiny)-naphthalene-6-sulfonic acid, sodium salt
TpeL  Clostridium perfringens toxin
TPEN  N,N,N',N'-tetrakis(2-pyridylmethyl)ethylenediamine
Tris  Tris(Hydroxymethyl)aminomethane
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDP</td>
<td>Uridine diphosphate</td>
</tr>
<tr>
<td>VcRTX</td>
<td>MARTX toxin from <em>Vibrio cholera</em></td>
</tr>
<tr>
<td>WT</td>
<td>Wild type</td>
</tr>
<tr>
<td>XAS</td>
<td>X-ray absorption spectroscopy</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1-1. Genetic organization of the *C. difficile* pathogenicity locus. ........................................... 5

Figure 1-2. General overview of TcdA and TcdB binding and entry into mammalian cells. ................. 8

Figure 1-3. Structure of the CROP domain. ......................................................................................... 11

Figure 1-4. Receptor, glycan and neutralizing antitoxin antibody binding sites on TcdA and TcdB. ........................................................................................................................................... 14

Figure 1-5. Domain organization of TcdA at neutral and acidic pH. ..................................................... 17

Figure 1-6. Structures of the VcRTX, TcdA, and TcdB CPDs. ......................................................... 20

Figure 1-7. Structure of the TcdB glucosyltransferase domain. ......................................................... 22

Figure 2-1 Alignment of the translocation domain of the large clostridial toxin family using ClustalX2.1 ................................................................................................................................. 38

Figure 2-2 Secondary structure prediction for the translocation domain of TcdB using JPRED3. ........................................................................................................................................... 39

Figure 2-3. Hydropathy analysis of the translocation domain. ............................................................. 40

Figure 2-4. High throughput mapping of the functional determinants in the translocation domain of TcdB. ........................................................................................................................................... 42

Figure 2-5. Characterization of defective purified TcdB mutants ......................................................... 44

Figure 2-6. Proposed model for TcdB translocation ............................................................................ 46

Figure 3-1. *In vitro* and *in vivo* toxicity of mutant TcdB. ..................................................................... 53

Figure 3-2. Characterization of mutant TcdB. ....................................................................................... 55

Figure 4-1. Structure of TcdA. ............................................................................................................... 66

Figure 4-2. Zinc is required for autoprocessing activity ........................................................................... 67
Figure 4-3. The delivery domain provides an extended scaffold for an $\alpha$-helical hydrophobic stretch involved in pore formation. ................................................................. 69

Figure 5-1. Identification of functional defects at the boundary of the CROP and translocation domain. .................................................................................................................................................................................. 80

Figure 5-2. Characterization of defective purified TcdB mutants Y1824K and N1839K......... 82

Figure 5-3. Characterization of receptor binding of defective mutants. ................................. 84

Figure 5-4. Characterization of TcdB C-terminal truncations. .............................................. 86

Figure 5-5. Cell surface binding activity of the CROP domains from both TcdA and TcdB. .... 88

Figure 5-6. Binding model of CSPG4 and bezlotoxumab to TcdB. ........................................... 90

Figure 5-7. Alignment of short and long repeats within the CROP domain of TcdB. .............. 91

Figure 6-1. The link between diphtheria, TcdA and TcdB. ...................................................... 97

Figure 6-2. Structure of TcdA, with emphasis on the translocation domain. ............................. 99
List of Tables

Table 1-1 Protein substrate of TcdA and TcdB* .......................................................... 24
List of Appendices

Appendix Figure 1. Validation of GC-enriched, codon-optimized TcdB protein from E.coli expression. ................................................................. 113

Appendix Figure 2. Defective TcdB mutants are folded and display a similar pH-induced unfolding by TNS fluorescence. ................................................................. 114

Appendix Figure 3. Alignment of the hydrophobic region of different forms/strains of TcdB. 115

Appendix Figure 4. Wall-eyed stereo view of representative electron density. ......................... 116

Appendix Figure 5. A truncated TcdA construct (TcdA1795) undergoes autoprocessing more efficiently than TcdA and TcdA1832. ................................................................. 116

Appendix Figure 6. Zinc is present in the APD active sites of TcdA........................................ 117

Appendix Figure 7. Topology of the TcdA delivery domain. Cylinders indicate α-helices, arrows represent β-strands ................................................................. 118

Appendix Figure 8. Full Western blots associated with Figures 3c and 3d............................... 119

Appendix Figure 9. Domain structures of TcdA and TcdB showing short repeats (green) and long repeats (yellow) in the CROP domains......................................................... 120

Appendix Figure 10. Defective TcdB mutants are folded properly.................................... 121

Appendix Table 1. Data collection and refinement statistics. .................................................. 122

Appendix Table 2. Alignment of enzyme domains to isolated domain structures................. 123

Appendix Table 3. Zinc to protein ratios determined by ICP-MS........................................ 123

Appendix Table 4. Primers used to generate the plasmids for this study......................... 124
Chapter 1

Introduction
1 Introduction

1.1 Clostridium difficile: Historical Perspective

Clostridium difficile is a gram-positive, spore-forming bacterium, first described by Ivan Hall and Elizabeth O’Toole in 1935 when they were investigating bacterial colonization from normal intestinal flora of newborn infants (1). It was originally named Bacillus difficilis due to the difficulty of isolating the bacteria from the culture (1). Even though it was isolated from normal intestinal commensal bacterium, the organism was found to be toxic to animals, which was hypothesized to be due to production of secreted toxins (1). Because there was no evidence showing that presence of C. difficile caused diseases in the newborn, however, it was considered as part of normal intestinal flora for the next forty years. It remained as non-pathogen until 1978 when C. difficile was identified as the primary cause of pseudomembranous colitis (PMC) (2,3). PMC is a severe inflammation of the colon; plaques, or pseudomembranes, composed of fibrin, mucous, necrotic epithelial and neutrophils are characteristic feature for PMC (4,5). With the introduction and usage of broad-spectrum antibiotics, especially clindamycin, reported PMC increased dramatically, which fueled the search for the cause of PMC (6,7). C. difficile was identified as the primary isolate from the feces of patients under clindamycin treatment; the toxic components from Hall and O’Toole’s isolate were also found linked to PMC(1,4), all indicating that C. difficile was the pathogen causing gastrointestinal disease following antibiotic therapy.

1.2 Clostridium difficile Pathogenicity

1.2.1 Clostridium difficile associated diseases

C. difficile is currently the leading cause of hospital acquired, antibiotic-associated infections in humans. Symptoms of C. difficile infections (CDI), often referred to as C. difficile associated disease (CDAD) can range from mild and recurrent diarrhea to life-threatening conditions such as pseudomembranous colitis and toxic megacolon (8). CDI is typically triggered by antibiotic treatment, which can alter the normal gut flora creating an environment that allows C. difficile spore to germinate, colonize and proliferate. Disease symptoms are caused mainly by two toxic proteins secreted by the bacteria, toxin A (TcdA) and toxin B (TcdB). C. difficile is considered the leading cause of nosocomial infections in hospitals, causing 29,000 death per year in the United States with a total cost of over $3 billion per year, placing a great burden on US health
care system (9,10). From 1900s to 2000s, there was a dramatic increase in rate of CDI and number of CDI related deaths, due to the emergence of virulent strain NAP1/027 (ribotype 027), which produces higher level of toxins (11), has a higher rate of sporulation (12) and increased antibiotic resistance (13). As a result, *C. difficile* was classified as an ‘urgent antibiotic resistance threat’ by US Centers for Disease Control and Prevention in 2013 that requires immediate action (14).

### 1.2.2 *Clostridium difficile* lifecycle

Transmission of *C. difficile* is via oral-fecal route primarily of the spore form. Spores are dormant cells, highly resistant to harsh environment conditions like acidity of stomach, antimicrobials or disinfectants (15). As most antimicrobial compounds specifically target active cells, spores ensure that *C. difficile* can survive the immune systems and antibiotic treatments. As *C. difficile* is obligate anaerobic bacteria that is not able to survive in the oxygenated environment, they have to transmit to hosts in the spore form (16). In the stomach, vegetative cells are usually eliminated by the acidic environment, with only spores persisting and passaging through stomach. Germination of spores occurs in the duodenum, triggered by primary bile acids from the liver by the germinant receptor CspC (15,17). Following germination, spores outgrow and colonize in the intestine. In the caecum and colon, *C. difficile* bacteria colonize, produce toxins and begin producing spores again for excretion (15). Regulation of colonization and toxin production is closely related to host microbiota and metabolome (17,18). The first step of pathogenesis is colonization where *C. difficile* can adhere to the mucus layer and degrade it by secreting mucolytic enzymes like protease Cwp84 for penetration (19). The second stage for pathogenesis is toxin production. Although the environmental signal that regulates toxin production is not well understood, studies suggest that toxin production may be induced by stress like sub-inhibitory level of antibiotic treatment (20) or catabolite repression (21). Additionally, *C. difficile* has exceptional ability to acquire antimicrobial resistance to increase a chance of survival (22,23). *C. difficile* possesses a highly mobile genome composed of transposons, insertion sequences and phages, which are commonly known to contribute to antibiotic resistance (22,24). Increased resistance of *C. difficile* towards common used antibiotics like vancomycin, metronidazole and fidaxomicin has been reported (25-28). As mentioned above, the major virulence factors are the two large clostridial toxins, TcdA and TcdB.
1.2.3 Regulation of toxin production

Several factors contribute to the virulence of *C. difficile*. Toxin A and B are the major virulence factors causing the symptoms of CDI. The genes encoding toxin A and B are located in a pathogenicity locus (PaLoc) of 19.6kb in the bacteria genome (29). The PaLoc in the most pathogenic strain encodes the two toxins (A and B) and three other proteins for regulation of toxin production and release (TcdR, TcdE and TcdC) (29) (Fig.1-1). Non-toxigenic strains lack the PaLoc, but can convert from non-toxigenic to toxigenic by acquiring the locus through horizontal gene transfer (30). Deletions, insertions, and mutations in the PaLoc coding region are used to identify and define different naturally occurring toxino-types, which are compared to reference strain VIP10463 (31). Over 30 toxino-types have been identified, which are different from VPI10463 in toxin activity (32). TcdR is a member of the ECF family of RNA polymerase sigma factor and is responsible for positive regulation of TcdA and TcdB expression (33). TcdC appears to be an anti-sigma factor that represses toxin expression (34,35). There is a non-sense mutation within the tcdC gene of the supervirulent strain ribotype 027, leading to the inactivation of the depression function of toxin expression that was speculated to be associated with increased level of toxin production and virulence of this strain (11). Despite many studies focusing on TcdC, the function of TcdC remained controversial due to conflicting reports. The role of TcdE also remains unclear. It is homologous to bacteriophage holin proteins, which are responsible for secretion of progeny phages from infected host cells (36). The fact that TcdA and TcdB does not seem to have any recognizable secretion signal and no bacterial lysis is required for toxin release led to the speculation that *C. difficile* export TcdA and TcdB through a non-classic secretion pathway possibly facilitated by TcdE (37,38). Various environmental stimuli are also involved in regulation of toxin production. Branched amino acids (proline or cysteine) (39), glucose and other metabolizable carbon sources (21,40), butanol and biotin in the local environment all repress toxin expression (41-43), whereas short chain fatty acids (butyric acids), high temperature and sub-inhibitory concentrations of antimicrobials seem to stimulate toxin expression (37,41). Lastly, growth signals and cell density also contribute to toxin regulation. TcdA and TcdB are expressed when the bacterial enter stationary phase, possibly due to starvation (44,45). Overall, toxin regulation is closely related to the local gut environment and metabolic state of the bacterium.
Figure 1-1. Genetic organization of the *C. difficile* pathogenicity locus.

Schematic representation of the pathogenicity locus (Paloc). Toxin-encoding genes, *tcdA* and *tcdB*, are indicated by blue arrows, regulatory genes are shown in green (*tcdR*) or red (*tcdC*), and holing-encoding gene is shown in light green. The direction of the arrows reflects the direction of transcription. TcdR positively regulates expression of *tcdA* and *tcdB* and *tcdR* itself. TcdC is an anti-sigma factor that negatively regulate expression. TcdE is a putative holing protein that involves in secretion of toxins.

1.3 *Clostridium difficile* toxins

1.3.1 Role of TcdA and TcdB in CDI

Even though several factors are involved in *C. difficile* pathogenesis including adherence, sporulation, colonization and growth, TcdA and TcdB are responsible for causing the symptoms of CDI. TcdA and TcdB, with a molecular weight of 308 and 270 kDa, respectively, belong to a family of large clostridial toxins (LCTs). LCTs are a group of homologous, high molecular-weight proteins that, in addition to TcdA and TcdB include the *Clostridium sordellii* lethal (TcsL) and hemorrhagic (TcsH) toxins, *Clostridium novyi* α-toxin (TcnA), and *Clostridium perfringens* toxin (TpeL). They share sequence identity, ranging from 26-76% and possess similar structural architecture. LCTs are glycosyltranferases that modify and inactivate small GTPases of Rho and Ras proteins, leading to disruption of cell integrity and function. Following secretion, TcdA and TcdB can bind and enter the colonic epithelium to cause inflammatory response, tight junction disruption, mucosal and tissue damage, epithelial cell death, and fluid secretion (46,47). The specific roles of TcdA and TcdB in pathogenesis has been under debate over the years. Initially, TcdA was thought to be the more potent enterotoxin due to the fact that
administration of purified TcdA in animal intestines caused increasing secretion, mucosa damage and inflammation while TcdB failed to achieve the same effects in the ileal-loop model in rodents (46-48). However, TcdB was able to cause cell death if there was prior intestinal damage or a sub-lethal dose of TcdA was added (46). Interestingly, TcdB is 100- to 1000-fold more potent than TcdA in intoxicating epithelial cells in cell culture. Thus, it was proposed that the two toxins might work synergistically with TcdA acting first to disrupt epithelial layer and allow TcdB to enter and cause further tissue damage (49,50). The role of TcdA and TcdB was challenged with the identification of C. difficile strain that only secretes TcdB but not TcdA (A⁻ B⁺); the A'B⁺ strain can still cause the same hallmark symptoms of CDI in humans (49), suggesting that TcdB alone is capable and sufficient for causing CDI. Role of TcdA and TcdB in disease was further investigated by generating isogenic C. difficile strains with specific toxin region deletions in animal infection models. Studies showed that the mutant strain producing only TcdB can cause fulminant disease in infected animals similar to the wild type strain with both toxins; whereas, mutant producing only TcdA is less virulent compared to wild type or TcdB-only mutant strains (51,52). Consistent with previous observations that naturally occurring A'B⁻ C. difficile strains are non-pathogenic, they showed that the isogenic double mutant which produces neither TcdA nor TcdB was completely avirulent in infected animals, indicating disease are strictly induced by TcdA and TcdB (51). Although it is still unclear whether there is major difference between the function of TcdA and TcdB in disease pathogenesis, these studies suggest that while both TcdA and TcdB are capable of inducing immune and inflammatory responses, TcdB seems to be the key virulence factor in causing fulminant disease (53).

1.3.2 Structure and function of TcdA and TcdB

TcdA and TcdB are large glucosylating proteins that have 2,710 and 2,366 amino acids, respectively; they share very high sequence identity to 48% and are composed of similar four functional domain structure (Fig.1-2).

TcdA and TcdB belong to the traditional AB toxins, where A subunit is the enzymatic domain and B subunit is responsible for delivery of enzymatic A subunit into the cytosol. The A-subunit of C. difficile toxins is the N-terminal glucosyltransferase domain (GTD) that modifies Rho/Ras proteins by glucosylation. For C. difficile toxins, the B-subunit is expanded into three sub domains: an autoprotease domain (APD), a pore-formation and translocation domain (TD), and
the C-terminal combined repetitive oligopeptide (CROP) domain (Fig. 1-2). The toxic protein utilizes the four domains to enter and intoxicate host cells by a four-step mechanism: 1) receptor binding and endocytosis, 2) pore formation and translocation, 3) autoprocessing and release of the GTD domain in the cytosol, and 4) inactivation of host GTPases by enzymatic GTD domain (Fig 1-2). Despite not having a full-length structure for TcdA or TcdB, structures for GTD, APD and CROPs have been solved and extensively studied in the past decade. (54-59). The structure of TcdA and TcdB shown by electron microscopy and X-ray crystallography analysis suggests that the overall organization and structure of the two homologous toxins is similar.
**Figure 1-2. General overview of TcdA and TcdB binding and entry into mammalian cells.**

(a) Schematic of TcdA and TcdB, with the glucosyltransferase domain (GTD) in red, the autoprotease domain (APD) in blue, the translocation domain in yellow, the hydrophobic region in orange and the C-terminal binding repetitive region (CROP) in green. (b) General overview of TcdA and TcdB binding and entry into mammalian cells. Uptake begins when TcdA and TcdB bind to cell surface receptors. There are no identified receptors for TcdA, while there are three identified receptors for TcdB (CSPG4, FZD, PVRL3). After binding to cells, TcdA is internalized into membrane-bound vesicles through a dynamin-dependent and clathrin-independent pathway, while TcdB is internalized through a clathrin mediated and dynamin-dependent pathway. These vesicles later fuse with early endosomes, and in response to the low
pH of endosomes, the hydrophobic region unwraps from the translocation domain and inserts into the endosomal membrane, forming a double helical hairpin pore. At low doses of TcdA/TcdB, the GTD and APD translocate through the pore, where the APD proteolytically cleaves and releases the GTD in response to intracellular inositol hexakisphosphate (Inositol-P₆). The free GTD glucysolyates Rho family GTPases, leading first to cytopathic effects, and later, cytotoxic effects. At high doses of TcdA/TcdB, a necrotic pathway is stimulated and is independent of GTD.
1.3.2.1 Cellular binding and uptake

To initiate entry into target cells, intracellular acting toxins must first bind a receptor or receptors on the host cell surface. Typically, binding is carried out by a dedicated functional domain that is part of the B domain of the toxin and interacts with a host receptor that is either a protein or glycolipid. For TcdA and TcdB this simple paradigm does not appear to apply.

1.3.2.1.1 The CROP domain

The receptor binding domain was historically assumed to be the C-terminal region of TcdA and TcdB, spanning residues 1832-2710 and 1834-2366, respectively (60-62). This region is composed of characteristic repetitive sequences thus termed combined repetitive oligopeptide domain (CROP domain) (Fig 1-3A) (61,62). The CROP domain is composed of repeating units of 19-24 amino acid short repeats (SRs) interspersed with four to seven long repeats (LRs) of 30-31 residues (61,62). TcdA CROP domain consists of 32 SRs and 7 LRs; TcdB CROP domain is considerably shorter, with 19 SRs and 4LRs. The crystal structure of a very C-terminal fragment of TcdA CROP domain reveals a β-solenoid structure formed by four SRs and one LR (58) (Fig 1-3A). Each repeat is comprised of a β-hairpin followed by a loop; the long repeat flanked by short repeats introduces a kink that disrupts the screw-like structure formed by short repeats. Based on the structure of this repeating unit, models of complete TcdA and TcdB CROP domains were constructed (Fig 1-3B). Although TcdA and TcdB share sequence homology and similar solenoid fold, they differ in length and overall organization of their repeating units. TcdA CROP shows an extended S-shaped structure, while a shorter and more rigid horseshoe-like is illustrated for TcdB CROP (Fig 1-3B)(58). The predicted tertiary structures were later confirmed by electron microscopy studies.
Figure 1-3. Structure of the CROP domain.

(A) The TcdA CROPs domain is made up of 32 short repeats (SRs) with seven interspersed long repeats (LRs), which are represented by green and blue boxes, respectively. The TcdB CROPs domain consists of 19 SRs and 4 LRs. (B) Models of the TcdA and TcdB C-terminal repeats were constructed based on the structure of TcdA CROPs f1 (58,59). The models are colored as in (A). Proposed CROP domain of TcdA and TcdB modified from reference (63).
1.3.2.1.2 Role of CROP domain in C. difficile toxins

Truncation studies have shed light on the role (or lack thereof) of the CROP domains in receptor binding. Barroso and co-workers showed that C-terminal truncations of TcdB tested in crude lysate retained some level of cytotoxicity, suggesting an alternative binding region might exist outside CROPs (64). More recently, a newly identified member of the LCT family called TpeL from C. perfringens was identified and found to be toxic to mammalian cells despite lacking a CROP domain (65). The role of CROP domain has been investigated further by removing the CROPs from TcdA and TcdB. TcdA devoid of the CROP region had reduced potency but was still cytopathic, suggesting that CROP region may be dispensable for toxin uptake, and that a second receptor binding region may be present upstream of the CROP (66). Likewise, a series of TcdB truncations suggest that the majority of CROP domain can be removed, and that 1-1493 may contain the minimal region required for cellular uptake (67). Together these studies provide compelling evidence that the CROP domain is not the sole receptor binding domain for both TcdA and TcdB.

1.3.2.1.3 TcdA receptors

Early evidence for the existence and location of a receptor-binding moiety for these toxins came from the recognition that the C-terminal repeating structure of TcdA was homologous to the carbohydrate binding region of streptococcal glycosyltransferases (68). Based on this homology, the CROP region of TcdA was proposed to be the receptor-binding domain of TcdA. Further, antibodies raised to recombinant CROPs can block the ability of TcdA CROPs by competing with full length TcdA for cell-surface binding (69,70). Consistent with this, TcdA was shown to interact with carbohydrate structures sharing a core structure of β-Gal-(1,4)-β-GlcNAc (71,72). Crystal structure of TcdA CROP identified the sugar binding site located at the kinked region between SR and LR of TcdA, suggesting a multivalent glycan binding model for TcdA (59). Despite all the data indicating carbohydrates binding to TcdA (61,62,73,74), the physiological relevance of these interactions is not known. Rabbit sucrose-isomaltase was first reported as a TcdA receptor, however cells and tissues that did not express this receptor were found to be sensitive to TcdA, including the human colonic epithelium (75). Later human gp96 was suggested as another TcdA receptor, although knock down of gp96 did not confer 100% protection against TcdA, suggesting alternative receptors for TcdA (76).
1.3.2.1.4 TcdB receptors

Whereas the specific receptors for TcdA remain unknown, application of genome wide screenings led to the discovery of three TcdB specific receptors (Fig 1-4). A whole-genome human shRNA Amir library screen identified chondroitin sulfate proteoglycan 4 (CSPG4) as the first TcdB receptor (77). At picomolar concentrations of TcdB, CSPG4-knockout cells conferred protection to TcdB induced cell rounding and apoptotic death. Soon after identification of CSPG4, LaFrance and co-workers used gene-trap insertional mutagenesis to isolate poliovirus receptor-like 3 (PVRL3) as a cellular receptor that mediates necrosis cell death, triggered by nanomolar doses of TcdB (78). Contrary to CSPG4, knock down of PVRL3 showed no protection of cells to TcdB-induced cytopathic cell rounding and apoptotic cell death, suggesting that PVRL3 is involved in a distinct downstream pathway. Most recently, a third receptor for TcdB was identified, highlighting the complexity of TcdB binding and entry. A CRISPR-Cas9-mediated genome-wide screen identified members of the Wnt receptor frizzled family (FZDs) as receptors for TcdB (79). Introduction of ectopic expression of FZDs was able to restore the same toxin mediated cytopathic effects of TcdB as in CSPG4 knock out cell lines, suggesting FZDs function as an alternative receptor to CSPG4, with both mediating the same downstream cytopathic and apoptotic cell death under picomolar toxin concentration. Direct interaction of either PVRL3 or FZDs with TcdB fragments devoid of the CROP indicates that both receptors bind outside of CROP domain (78,79). Pull-down studies suggest that PVRL3 binds to TcdB in the 1372-1493 region (67), and that FZDs bind to TcdB in the 1500-1830 region (79). Interestingly, CSPG4 is able to associate with truncated fragments of TcdB 1500-2366, but not to TcdB 1852-2366, suggesting that the binding site for CSPG4 lies in somewhere between amino acid 1500-1852 (77). Competition assays show that FZDs and CSPG4 bind independent of each other, suggesting they might have distinct, non-overlapped binding site (79).

1.3.2.1.5 CROP-targeted neutralizing antibodies prevent receptor-binding

Although the CROP domain appears to be dispensable for binding and function of TcdB, antitoxin antibodies that target the CROP domain can still neutralize the cytotoxic effects of both TcdA and TcdB. Two human monoclonal antibodies, actoxumab and bezlotoxumab, have demonstrated reduction in the rate of C. difficile infection (CDI) recurrence in clinical trials (80-83). These two antibodies are shown to neutralize TcdA and TcdB respectively, by blocking toxin binding to mammalian cells; interestingly, both antibodies bind to the CROP domain of
their respective toxins (84,85) (Fig 1-4). The antitoxin for TcdA, actoxumab, has two spatially distinct binding sites identified within TcdA CROP domain; epitope 1 located in the middle of the CROP domain (residues 2162-2189) and epitope 2 is closer to the C-terminus of the CROP (residues 2410-2437) (Fig 1-4) (85).

Figure 1-4. Receptor, glycan and neutralizing antitoxin antibody binding sites on TcdA and TcdB.

*(top)* There are no physiologically-relevant identified receptors for TcdA. Two glycan core structures have been identified that bind to the CROP region with high affinity: Gal β-(1,4)-GlcNAc and α -Gal-(1,3)- β-Gal- (1,4) -β- GlcNAc. There are two actoxumab binding sites within the TcdA CROP: residues 2161-2189 and 2410-2437. *(bottom)* Three receptors have been identified that bind to TcdB and their putative binding regions were determined from functional assays. The majority of C-terminal region of delivery domain is proposed to contribute to receptor binding, shown in light green. CSPG4 is postulated to bind in the region encompassing both the translocation domain and the CROP domain at residues 1500-1852, FZD to residues 1500-1830 and PVRL3 to 1372-1493. The neutralizing antibody bezlotoxumab binds to two epitopes within N-terminal of TcdB CROP domain, at residues 1902-1947 and 2033-2092.

At high antibody: toxin concentrations, multiple antibodies can bind to TcdA via both epitopes, forming large immunocomplexes. Electron microscopy studies indicate that actoxumab-Fab binding to both epitopes alters protein conformation within TcdA, shown as a loss of resolution of the GTD, APD and the translocation domain (85). Even though the identified epitopes overlap with potential hydrophobic binding sites for TcdA, actoxumab only occupies two out of the seven TcdA carbohydrate binding pockets, suggesting that neutralization is not entirely due to direct blocking of carbohydrate binding sites (85). The possibility that conformational changes caused by actoxumab prevents TcdA binding to receptors cannot be ruled out. Bezlotoxumab,
the neutralizing antibody for TcdB, was recently approved by US Food and Drug Administration (FDA) for treating recurrence of CDI. Biophysical and structural data indicate that bezlotoxumab binds to two homologous sites within N-terminal region of the TcdB CROP domain (residues 1834-2101) (84). While both bezlotoxumab and actoxumab have two epitopes located within the CROP domain, the relative location of two epitopes is quite different; unlike actoxumab, a single molecule of bezlotoxumab can bind simultaneously to two epitopes in close proximity via Fabs (84). The one-to-one binding ratio suggests that bezlotoxumab can directly neutralize TcdB without forming an immunocomplex (84).

1.3.2.2 Endocytosis

A central step in the intoxication pathway of translocating toxins that follows cell-surface engagement is receptor-mediated endocytosis into acidified vesicles. Interestingly, TcdA and TcdB appear to use different pathways to enter cells possibly as a result of using different receptors for binding. TcdA internalization was initially shown to be accompanied by clathrin coated pits (86,87), while TcdB internalization was first reported to be clathrin-mediated, and dynamin dependent (88). Addition of the dynamin inhibitor dynasore or the clathrin-dependent pathway inhibitor chlorpromazine significantly delayed TcdB-mediated intoxication (88). RNAi-mediated knock down of clathrin blocked endocytic uptake of TcdB, confirming the importance of the clathrin-mediated pathway in TcdB internalization (88). When a moderate reduction in toxin uptake was observed for TcdA treated with clathrin or dynamin inhibitor, it was suggestive of an alternative clathrin-independent pathway (89). Unlike full-length TcdA, the endocytic uptake of CROP-less TcdA was almost fully blocked by inhibiting dynamin, suggesting that an alternative binding region upstream of the CROP domain exists and it utilizes different endocytosis pathway independent of clathrin but dependent on dynamin. The mechanism of TcdA endocytosis was later elucidated by Chandrasekeran et al. to be dynamin-dependent but independent of clathrin or caveolae (90). Host factors from the caveolar pathway, protein kinase C and casein kinase substrate in neurons 2 (PACSIN2), are proposed to be essential for TcdA uptake. The implication of downstream cytotoxic effects and the role of CROP-dependent and – independent endocytic pathways remain to be studied.
1.3.2.3 Pore formation and translocation

Perhaps the most remarkable feature that translocating toxins perform during their uptake and entry is their escape from endosomes into the cytosol. Typically, in response to the low pH of endosomes, a specialized pore-formation/translocation domain undergoes a conformational change that leads to insertion of marginally hydrophobic segments – that were buried within the pre-pore structure – into the bounding membrane to create a pore that is permeant to ions. This stable pore is thought to serve as the conduit for the associated toxin enzyme domain(s) to reach the cytosol (91-93). Recent studies on the structure and function of the pore-formation/translocation domain of TcdA and TcdB highlight certain unexpected similarities and differences with other translocating toxins (94).

1.3.2.3.1 TcdA and TcdB undergo pH-dependent conformational changes that are required for toxicity

Like other translocating toxins that require acidic pH to escape endosomes, the toxicity of TcdA and TcdB can be blocked using agents that inhibit vacuolar ATPases, such as bafilomycin A1 and other lysosomotropic agents that prevent endosomal acidification (95). Ballard and co-workers demonstrated that acidic pH resulted in exposure of hydrophobic surfaces and changes to protease susceptibility, consistent with the idea that hydrophobic regions must become exposed to insert into membranes (95). Structural changes between neutral and acidic pH were later revealed by electron microscopy (EM) (56) (Fig. 1-5) and refined by fitting the high-resolution TcdA crystal structure to the EM map (94). At neutral pH, the toxin takes on a bilobed structure, with the translocation domain making contact with the GTD and CROP region. At acidic pH, the translocation domain becomes more elongated, and the contact points between the translocation domain and CROP are lost.
Figure 1-5. Domain organization of TcdA at neutral and acidic pH.

EM structures are colored by functional domain. GTD, yellow; CPD, blue; delivery domain, red; CROPs, green. The dashed line represents possible unfolding of the TcdA GTD N-terminus at low pH. EM figure adapted from reference (96).
In light of the toxin’s dependence on acid pH for toxicity, it was postulated that TcdA and TcdB’s strategy to intoxicate might be similar to other acid pH dependent toxins, namely anthrax (92), diphtheria (91) and botulinum toxin (93). In response to low pH, TcdA and TcdB were also shown to undergo pH induced structural changes that trigger membrane insertion and pore formation of TcdA and TcdB (97,98). To evaluate pore formation, the authors measured the ability of TcdA and TcdB to release rubidium from pre-loaded cells by mimicking the endosomal environment at the cell surface. Both toxins require acidic pH to cause significant rubidium release from cells, which is interpreted as the toxin’s ability to insert into membranes and form pores; additionally, TcdA requires cholesterol-enriched membranes for insertion (98). The authors also noted that pore formation was accompanied by ion channel formation, which has historically been indicative of pore formation for other translocating toxins (99-101).

1.3.2.3.2 Structural determinants for pore formation

Electrophysiology has proven to be an invaluable tool for studying pore formation and translocation of toxins. Most commonly, the planar lipid bilayer (PLB) technique is used, which involves forming an artificial membrane across an aperture separating two compartments, applying voltage across the bilayer and measuring the resulting current (102). All toxins that translocate out of endosomes form stable ion-conductive channels that are pH-dependent, with most having conductances <100 pS (99,101,103-106). TcdA and TcdB, on the other hand, do not appear to form stable ion-conductive pores; instead, both toxins exhibit characteristic “flickering” electrophysiological behavior, with large conductances of up to ~1-2 nS and lifetimes of ~several milliseconds (97,98,107). While it is still unclear whether flickering is indicative of toxin-membrane interaction, or more specifically, membrane insertion, pore formation and/or translocation, this behavior does appear to correlate well with rubidium release experiments. Point mutants that attenuate pore-formation (as measured in standard rubidium experiments) show greatly reduced or absent electrophysiological activity.

In light of the ability of TcdA and TcdB to form pores as a result of low pH, many have sought to identify structural determinants for membrane insertion, pore formation and translocation. Genisyuerek et al was the first group to describe a minimal pore-forming region of TcdB covering amino acid 830 to 990 (108). They used deletion studies to show that N-terminal region including GTD and APD is not involved in pore formation while a fragment of 1-990 is capable
of inducing pore formation in artificial membranes and lipid vesicles, thus narrowing down the pore-forming region to TcdB 830-990. Additionally, a pair of amino acids glutamate-970 and glutamate-976 was proposed as pH sensor that is crucial for pore-formation. It is acknowledged that, however, the region seems too small to be able to translocate the enzyme domains into cytosol, indicating there must be additional region involved in the process. Once the primary sequences for TcdA and TcdB were determined, it was postulated that the most hydrophobic region of the toxins (i.e., amino acid 958-1130 in TcdA; 956-1128 in TcdB) in the translocation domain was implicated somehow in membrane insertion, pore-formation and translocation (56,61,73,97,98). Unfortunately, there was no further information regarding the determinants for pore formation and translocation.

1.3.2.4 Autoprocessing (APD)

All clostridial toxins possess a unique autoproteolysis mechanism to release toxic domain in the cytosol. In 2003, Pfeifer first identified that only the N-terminal GTD was found in the cytosol while the remaining toxin was localized in the endosome, suggesting the toxin is proteolytically cleaved in the cytosol (109). Rupnik et al showed that the cleavage reaction in vitro specifically requires the condition at neutral pH and presence of host cell cytosol (110). Subsequently, inositol hexakisphosphate (InsP6) was demonstrated to activate the autoproteolytic cleavage (111). The concentration of InsP6 in mammalian cells is approximately 10-60 µM (112), which is sufficient for effective autoproteocleavage. The enzyme responsible for auto-cleavage is located at C-terminal of GTD; it is a cysteine protease that is homologous to the cysteine protease domain of MARTX toxins (VcRTX), also induced by InsP6 (113). The cysteine protease domain covers residues 543-769 in TcdA and 543-767 in TcdB; the cleavage occurs at a highly-conserved Leucine residue at Leu542 in TcdA and Leu 543 in TcdB and releases the full GTD into cytosol (110,114). There are three conserved residues (Cys 698, His 653 and Asp 587 in TcdB; Cys 700, His 655 and Asp 589 in TcdA) at the active site; mutating any of the residues results in the complete inhibition of protease activity (57,115) (Fig. 1-6). It is suggested to rename the domain from cysteine protease (CPD) to autoprotease (APD) because the recent discovery of a zinc ion at the active site that challenged the previously termed cysteine protease domain (94). The zinc seems to be essential for protease activity but the molecular mechanism is unclear.
Figure 1-6. Structures of the VcRTX, TcdA, and TcdB CPDs.

The crystal structures of the CPDs are shown as ribbon diagrams. InsP6 and the catalytic residues are shown as sticks (red and orange, respectively). The β-flap, shown in purple, separates the InsP6 binding pocket from the catalytic site. The N-terminus (dark blue) wraps around the protein such that the N-terminal cleavage site is near the catalytic residues. Figure adapted from reference (63).

Autoproteolysis is induced and greatly enhanced by InsP6. InsP6, a highly negatively charged molecular, was found binding to APD in a highly positively charged pocket by a 1:1 ratio in crystal structure analysis (116). Binding of InsP6 can cause significant conformational changes to the structure, which is thought to be essential for activation (57). The InsP6 binding site is separate from the active site by “β-flap” structure formed by a three-stranded β sheet (57) (Fig 1-6). Following binding of InsP6, there is a major allosteric change transduced by the β-flap to the active site (54). One outcome of the structural change is to increase the number of positively charged residues at the active site.

Although autoprotease domains of TcdA and TcdB are very similar, the cleavage efficiency for these two toxins varies. In vitro, TcdB is more sensitive towards InsP6 activated cleavage than TcdA (114). TcdA autoprocessing seems to be repressed by the presence of TcdA CROP domain by intermolecular interaction with N-terminal of holotoxin, which disappears upon protein unfolding triggered by acidification (94,117,118). The efficiency of autoprocessing activity has
an impact on virulence; *C. difficile* toxins with inactivated-APD were shown to have reduced toxicity in infected host (119). However, APD-inactivated mutants of TcdA and TcdB are still capable of glucosylating small GTPases and causing cytopathic effects in cells (120). When the toxins are unable to release GTD into the cytosol freely, GTD can still contact and inactivate substrates, although at a reduced rate. As a result, cytotoxicity is reduced but not abolished. It is probably due to the fact that the one of substrates for TcdA and TcdB is Rac-1, which is recycled and activated at the endosome, so GTD remained tethered to the endosome could still get access to some of the substrates (121). In conclusion, autoprocessing is important for toxin potency and virulence by mediating the release of GTD into cytosol and the rate of substrate inactivation, whereas it is not essential for the cytotoxicity of TcdA and TcdB.

1.3.2.5 Biological activity of *C. difficile* toxins

1.3.2.5.1 Structure and function of glucosyltransferase domain (GTD)

Following autoprocessing, the enzymatic domain is released into the cytosol to act on host targets. The enzymatic domain for TcdA and TcdB is the encoded N-terminal GTD, which is a glucosyltransferase that inactivates small GTPases from the Rho family (122,123). The glucosyltransferase domain (GTD) locates at the N-terminal of the toxin covering from amino acids 1 to 543 in TcdB and 1 to 542 in TcdA (Fig 1-7).
Figure 1-7. Structure of the TcdB glucosyltransferase domain.

(A) The core GT-A fold is shown in orange, and the α-helical additions are shown in red. The N-terminal four-helix bundle makes up the MLD. (B) Close up view of the catalytic core. UDP, glucose, and a manganese ion are bound on the surface of the core GT-A fold. The binding pocket is overlayed by a loop comprising residues 517–523. Some of the residues involved in coordinating UDP, Glucose, and Mn$^{2+}$ are shown as sticks. (C) Surface view of the TcdB GTD as shown in (A). Glu449, Arg455, Asp461, Lys463, and Glu472, residues that have been shown to be involved in GTPase binding, are colored cyan. The GTD structure figure adapted from reference (63).
The crystal structure of GTD domains from both toxins help better understand the structural features and related mechanisms of the enzyme. The enzymatic core of the glucosyltransferase is composed of parallel strands; the peripheral part is mainly helices (Fig. 1-7). The overall topology looks very similar to glucosyltransferase type A (GT-A) family, which typically adopts an α/β/α sandwich called Rossman-like fold (55,124). Shown in Figure 1-7, the very N-terminal end of the domain of amino acids 1 to 90 form a four-helix bundle that contains basic amino acid residues, which is suggested to be responsible for plasma membrane binding, therefore referred to as membrane localization domain (MLD) (125). Some of the key residues involved in enzymatic activity within GTD domain are highly conserved through LCTs. Within the enzyme core, there is a DXD motif (D285/D287 in TcdA and D286/D288 in TcdB) that is essential for coordination of binding of the manganese cofactor and UDP-glucose (126,127). The first aspartic acid binds Mn\(^{2+}\) directly; the second interacts with Mn\(^{2+}\) through a water molecule (Fig. 1-7). The first aspartic acid also coordinates with 3'-hydroxyl group of the UDP-ribose and the glucose. Binding of UDP-glucose is stabilized by two tryptophan residues (W101 and W519 in TcdA and W102 and W520 in TcdB) through interacting with the uracil of UDP-glucose by aromatic stacking and the glycosidic oxygen (128,129). Mutations at these conserved sites greatly inhibit enzymatic activity of the toxins (128,129). These GTD inactivated mutants are frequently used as control for experiments. It is important to notice that enzymatically-inactivated point mutants might not achieve full inactivation of the toxicity; for example, GTD defective mutant D270N can reduce activity by approximately 500 to 1000 fold (129).

1.3.2.5.2 Rho modification by GTD

The main target for glucosyltransferase of TcdA and TcdB are Rho family proteins Rho, Rac and Cdc42 (130,131). There are still differences in the substrate specificity between TcdA and TcdB (Table 1); for example, TcdA is able to inhibit Rap while TcdB does not (96). Additionally, different isoforms of TcdA and TcdB developed different substrate specificity compared to the reference strain (Table 1). Mutagenesis studies showed that residues located near the sugar binding pocket might contribute to the substrate specificity (129). Several TcdB variants contain mutants similar to other LCTs at proposed substrate binding site that might enable binding of these TcdB variants to additional Rac and Ras instead of Rho GTPases by classical TcdB, which is suggested to complement the role of TcdA (55). There is no crystal structure of GTD domain with its substrate. Although molecular basis for substrate specificity is not fully understood,
identified mutants suggest that the alpha-helical sub-domains are essentially for substrate recognition (129).

**Table 1-1 Protein substrate of TcdA and TcdB**

<table>
<thead>
<tr>
<th>Toxin</th>
<th>Strain</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>TcdA</td>
<td>VPI10463(^a)</td>
<td>Rho, Rac, Cdc42, Rap</td>
</tr>
<tr>
<td>TcdB</td>
<td>VPI10463(^a)</td>
<td>Rho, Rac, Cdc42</td>
</tr>
<tr>
<td></td>
<td>NAP1/RT027(^b)</td>
<td>RhoA, Rac, Cdc42, Rap, R-Ras</td>
</tr>
</tbody>
</table>

* Data is based on references 130, 132, 133.
  a. Reference strain
  b. Epidemic strain

Rho proteins are small 21-25 kDa GTP binding proteins that act as molecular switch for many signaling processes involving actin cytoskeleton and motile processes (132). Regulation of Rho proteins is through a GTPase cycle; GTP-bound form is active and GDP-bound form is inactive (133). Regulation of GTPases is mediated by three types of proteins. Guanine nucleotide exchange proteins (GEFs) activate the GTPase by exchange GDP with GTP (134). The active GTPase is terminated by GTPase activating proteins (GAPs) through GTP hydrolysis (133). The last but not the least, guanine dissociation inhibitors (GDIs) mediate extraction of GTPase from membranes and keep Rho family proteins in inactive form in the cytosol, which is important for Rho GTPase posttranslational modification and membrane association (135). The GTP bound active state of GTPases can interact with various downstream effectors to regulate cellular functions including organization of actin skeleton, which is key for maintaining the cell structure. The main target for TcdA and TcdB is the GDP-bound, inactive form of Rho protein (130). The inactivation occurs through attachment of a glucose from donor substrate UDP-glucose onto Rho proteins at conserved threonine 35/37 by GTD (130,131). The monoglucosylation reaction is formed by a \(\alpha\)-glycosidic configuration of UDP-glucose and Rho proteins, which is not reversible due to the absence of glucosidases in humans. The threonine residue is key in association of an \(\text{Mg}^{2+}\) ion required for GTP binding (136). The residue locates at the switch I region of the GTPase; binding of region by GTPase causes major conformational changes that allow activation and interaction with effector proteins (137,138). Toxin-induced glucosylation of
Rho proteins at Thr35/37 inhibit the interaction of Rho proteins with their effectors by steric hindrance, affecting downstream signaling (130). Furthermore, glucosylation also blocks interaction with regulatory proteins that locks Rho GTPase in the inactive form by preventing binding to GEFs and GAPs for activation and keep Rho GTPase from being extracted from cell membrane by GDI proteins (139,140). In summary, the main effect of toxin-induced glucosylation is to inhibit downstream signaling of GTPases and subsequently causes cytopathic and cytotoxic effects that ultimately links to disease symptoms.

1.3.3 Cellular effects of TcdA and TcdB

Inactivation of Rho proteins by TcdA and TcdB causes disruption of numerous cellular functions that contribute to characteristic symptoms of CDI like diarrhea, inflammation and colitis. Direct cellular effects following toxin treatment are epithelial cell death and disruption of tight junctions that result in the damaged barrier. Toxin can also trigger inflammatory responses by stimulating epithelial cells to release proinflammatory cytokines and neutrophil chemoattractants (141,142). The impaired barrier and acute inflammation will further cause neutrophil infiltration, increased intestinal and vascular permeability and fluid secretion (143). Entry of toxins and bacteria due to loss of the barrier can cause further tissue inflammation and damage, likely contributing to severe diseases symptoms including pseudomembranous colitis (141). We will discuss cellular effects of toxins and their relation to CDI symptoms below.

1.3.3.1 Cytopathic effects

The typical immediate cellular effect of TcdA and TcdB intoxication is the change in the cell morphology; cells tend to shrink and become rounded within 1 hour of the toxin treatment (144,145). The rounded phenotype on tissue cells is termed the cytopathic effect of C. difficile toxins. Inactivation of Rho proteins by C. difficile toxins directly causes the cell rounding phenotype. Because Rho GTPases (Rho, Rac and Cdc42) are involved in the regulation of the assembly and organization of actin cytoskeleton, cell morphology and polarity and cell-cell contacts, inactivation of Rho GTPase by toxin induced glucosylation results in loss in the cytoskeleton structure and disassembly of focal adhesions and tight junctions, represented by characteristic cell rounding effects (146-150). Disruption to the cell-cell contacts, cell adhesions and tight junctions can further damage the barrier and integrity of intestinal epithelium that will increase intestinal permeability and inflammation.
1.3.3.2 Cytotoxic/Apoptosis Effects

In addition to the early cell rounding effect, glucosylation of Rho GTPases also induces cell death by apoptotic mechanism, which was shown in intestinal epithelium cell lines and human intestinal tissue (119,151-154). In tissue culture, cell death is usually observed within 12-48 hours of intoxication, which is termed cytotoxic effect. Cytotoxic cell death exhibits typical effects of apoptosis: cell shrinkage, caspase activation and DNA fragmentation (151,155-159). Studies showed that induction of apoptosis is strictly dependent on glucosylation; the GTD-inactive mutants are incapable inducing apoptosis in vitro (151,160). Toxin induces activation of executioner caspase 3 and 7 and triggers apoptosis (157-159,161,162). Apoptosis is not solely due to redistribution of the actin cytoskeleton because it was shown that Latrunculin B, which depolymerizes actin filaments, did not induce apoptosis (160). Inactivation of Rac1 is essential for activation of caspase 3. Interestingly, selective inactivation of Rac1 showed different apoptosis induced pathways when compared to global inactivation of Rho proteins by TcdB, suggesting that other Rho proteins might also involve in the process (163).

1.3.3.3 Glucosylation-independent necrosis effects

TcdB has also been shown to induce cell death at a high molar concentration through a glucosylation-independent manner (164). While TcdA and TcdB can induce apoptosis in a glucosylation-dependent mechanism, very recently the Lacy group reported that at a high concentration (1,000 times more than concentration required for cytopathic and cytotoxic effects), TcdB is able to cause necrotic cell death, independent of either autoprocessing or glucosyltransferase activities (164,165). Under this condition, treated cells exhibit the hallmark feature of necrosis including rapid ATP depletion, loss of membrane integrity, lactate dehydrogenases (LDH) release and chromatin condensation within 1-2 hours of intoxication (164-166). The necrosis effect does not trigger caspase 3/7 activation; instead, necrosis is caused by activation of NADPH oxidase (NOX) complex on endosomes and subsequent aberrant production of reactive oxygen species (ROS) (165,166). High level of ROS production can trigger DNA damage, protein oxidation, lipid peroxidation and mitochondrial dysfunction, leading to necrosis (167-169). Enzymatically inactive mutants of TcdB were unable to induce ROS production and necrosis, whereas GTD domain is still required (166). Interestingly, the same effect is not observed in TcdA at all concentrations tested (94). The necrosis cell death seems to be specific to TcdB and independent of GTD or APD activity, which may explain why
variant strains that express either only TcdB (TcdA⁻TcdB⁺) or both TcdA and TcdB (TcdA⁺TcdB⁺) are more pathogenic in causing colonic epithelial damage than strains only expressing TcdA (TcdA⁺TcdB⁻) in infection models (53). Whether the TcdB-induced necrosis can contribute to severe gut damage observed in patients is unknown; the mechanism of necrosis and its relevance in CDI remain to be studied.

1.3.3.4 Toxin induced inflammation and its role in diseases

Inflammation is an essential factor leading to the hallmark feature of CDI infections from diarrhea to colitis, and is currently under active investigation. In addition to the short-term toxin mediated cytopathic and cytotoxic effects causing damaged epithelial barrier and epithelial cell death, intoxication of TcdA and TcdB also promotes release of inflammatory mediators that triggers neutrophil recruitment to colonic mucosa, which is associated with fluid secretion, mucosal inflammation and damage in animal infection model, typically found in CDI (5,142,170). These cytokines induce release of IL-6, IL-8 and Interferon-γ among which IL-8 is the most potent (153,171-173). Neutrophil invasion into mucosa is also typical for CDI, which can lead to severe mucosal damage. Disruption of the protective epithelial barrier tends to create an opportunity for toxins to enter lamina propria, directly stimulating innate immune cells including monocytes and macrophages to release inflammatory mediators. Innate immune responses to toxins activate inflammasome and pyroptosis, also inducing strong inflammation (174,175). Overall, the innate immune responses towards TcdA and TcdB infection involve multiple players and cell types and are essential in mediating C. difficile-associated diarrhea and colitis. Inhibition of release of inflammatory mediators or activation of inflammasome was shown to effectively reduce TcdA and TcdB induced inflammation and colonic damage in vivo (174-177). Consistent with the in vivo data, the level of inflammatory cytokines is correlated with CDI severity in CDI patients (178-180). It is important to highlight that innate immune response is essential at early stages of infection to contain luminal bacterial and prevent systemic dissemination; however, prolonged and over-stimulated neutrophil invasion and inflammation will likely contribute to increased pathogenic CDI infections. Thus, the stage of infection needs to be taken into consideration when modulating these immune responses when treating CDI.
1.4 Existing Problems and Thesis Rational

A major aim of my research is to elucidate the molecular mechanism by which *C. difficile* toxins mediate their own binding and entry into mammalian cells. Although there have been extensive studies on overall disease pathogenesis and individual domains of *C. difficile* toxins, there are still overarching questions remaining, especially on the mechanism of toxin entry into the cytosol, which is the critical step for toxin function.

The initial step for intoxication is binding of the toxin to cell surface receptors. There have been several studies over many years trying to identify putative receptors for TcdA and TcdB; it was not until recently that three receptors have been identified for TcdB while the exact receptor for TcdA is still unknown. Recent identification of receptors has advanced our knowledge on cell surface binding for TcdB, however, it also highlights the complexity of TcdB binding and entry and raises more questions; for example, what is the advantage of the apparent multiple-receptor model? and, how do different receptors contribute to toxicity and affect downstream pathways?

The fact that all identified receptors seem to bind outside of CROP domain, which has been assumed to be the sole receptor binding domain, and that toxin devoid of CROP domain or partial translocation domain is still able to bind and intoxicate cells questions the function of CROP domain in receptor binding and underlies the ambiguity in the seemingly well-defined boundary between the translocation domain and receptor binding domain. To elucidate the importance of the boundary and role of the CROP domain, a mutagenesis and reductionist approach was utilized here to delineate functionally important regions for receptor binding in Chapter 5. Previous studies mainly focused on making C-terminal truncations in the delivery domain to identify minimal required region for translocation (108), however, few studies focused systematically on the role of specific residues within the CROP domain. By making C-terminal truncations in the CROP domain, I aimed to identify the minimal required region for CROP binding and its relevance to identified TcdB receptors.

Recent findings suggesting that CROP is not the sole receptor binding domain also question the traditionally defined translocation domain. While there have been extensive studies and structural information on individual domains of GTD, APD and CROP domain, there exists a gap in our knowledge of the mechanism by which *C. difficile* toxins cross membranes to enter cells. Structural studies of TcdA and TcdB have relied primarily on a reductionist approach,
resulting in the elucidation of high-resolution structures for the individual GTD, APD and CROP domains, however, the delivery domain and full-length toxin has been refractory to high-resolution structural analysis. Negative stain electron microscopy was used to obtain low-resolution structure of full-length TcdB (56), which revealed the overall topology of the four domains at both neutral and acidic pH (56). These studies, though lacking in molecular detail, confirm that major conformational changes result upon acidification, particularly in the translocation domain. Like most A-B toxins, the translocation domain of *C. difficile* toxins forms ion-conducting channels in artificial lipid bilayers at acidic pH (98). However, a functional understanding of the mechanisms by which translocation domain transitions from a soluble “pre-pore” conformation to a membrane-inserted pore conformation that translocates its enzymatic domain across membranes has been lacking.

Structurally, the translocation domains from toxins have been shown to adopt one of the two structural motifs in the membranes: a bundle of α-helices (diphtheria and botulinum) (181,182), or β-barrel (anthrax) (92). For LCTs, it is unknown which motif is used to cross membranes. It is expected from other toxins that membrane spanning pore is generally composed of hydrophobic stretches that are at sufficient length to span membranes. However, there is no evidence to indicate that LCTs form oligomers at any stage, which is typical for β-barrel pores. Identifying transmembrane segments in bacterial toxins that only insert into membranes at acidic pH is not straightforward using standard hydropathy plots, as the polarity of certain amino acids varies with pH. It has been suggested that residues 956-1128 may constitute the region that forms the pore domain due to hydrophobicity (56,98). We aimed to use a scanning mutagenesis approach to delineate the membrane-inserted domains of TcdB and conclusively define the regions within the translocation domain that insert into the membrane, which will be discussed in detail in Chapter 2.

I hypothesized that at low pH in the endosome, marginally hydrophobic segments rich in acidic residues insert into membranes as α-helices that associate through specific helix-helix interaction motifs to form pores that are lined with conserved features, which are essential for GTD unfolding and translocation. In Chapter 2, a comprehensive library of mutants across the entire translocation domain was used to identify the functional determinants of TcdB toxicity on cells. A panel of biochemical and biophysical assays were used to determine each of the determinants in membrane-insertion, pore-formation, and translocation. The difficulty to study the full-length
toxin and translocation domain through biochemical structural-function methods is the size of the protein, with translocation domain being the 1000-residue long and is the largest domain among the four domains. To best approach constructing mutations to a near 1000-residue domain, I prioritized amino acids based on their degree of conservation and hydrophobicity. Additionally, \textit{tcdB} gene is highly AT-rich, making it extremely difficult for DNA manipulation and expression in \textit{E.coli}. To tackle that problem, we generated a codon-optimized DNA sequence for TcdB with increased GC content to increase efficiency. Through the mutagenesis approach, I identified key residues essential for pore-formation; based on these key determinants I proposed a double-dagger model for TcdB entry into cytosol. In Chapter 3, the role of key pore-forming mutants was probed in detail by \textit{in vitro} and \textit{in vivo} studies. In Chapter 4, in collaboration with Dr. Lacy, we described the recently solved crystal structure of translocation domain from TcdA, which shed light on the unique structure of hydrophobic pore-forming region of \textit{C. difficile} toxins and helped further refine our model of translocation.

Our characterization of the translocation domain and CROP domain help to further clarify domain boundaries and functional determinants within for both pore formation, translocation and receptor binding.
Chapter 2

Translocation Domain Mutations Affecting Cellular Toxicity Identify the Clostridium difficile Toxin B Pore

The work presented in this chapter was published in Proc. Natl. Acad. Sci. U.S.A journal.


Contributions:
My contributions to this work included designing the experiments, generating and expressing constructs and test mutants for mutagenesis screening as well as most of the characterization of defective mutants. The hydropathy analysis and proposed model was performed and analyzed along with R. Melnyk. Planar lipid bilayer experiment was performed by R. Melnyk. M. Park assisted with mutant library construction. J.Tam, A. Auger and G. L. Beilhartz assisted with analytic tools. D.B. Lacy and R. Melnyk assisted with analyzing the data. R. Melnyk supervised the project.
2 Translocation Domain Mutations Affecting Cellular Toxicity Identify the Clostridium difficile Toxin B Pore

2.1 Overview

Disease associated with Clostridium difficile infection is caused by the actions of the homologous toxins TcdA and TcdB on colonic epithelial cells. Binding to target cells triggers toxin internalization into acidified vesicles, whereupon cryptic segments from within the 1050 amino acid translocation domain unfurl and insert into the bounding membrane creating a transmembrane passageway to the cytosol. Our current understanding of the mechanisms underlying pore-formation and the subsequent translocation of the upstream cytotoxic domain to the cytosol is limited by the lack of information available regarding the identity and architecture of the transmembrane pore. Here, through systematic perturbation of conserved sites within predicted membrane-insertion elements of the translocation domain, we uncovered highly sensitive residues – clustered between amino acids 1035 and 1107 – that when individually mutated reduced cellular toxicity by as much as >1000-fold. We demonstrate that defective variants are defined by impaired pore-formation in planar lipid bilayers and biological membranes, resulting in an inability to intoxicate cells through either apoptotic or necrotic pathways. These findings along with the unexpected similarities uncovered between the pore-forming “hotspots” of TcdB and the well-characterized α-helical diphtheria toxin translocation domain provide the first insights into the structure and mechanism of formation of the translocation pore for this important class of pathogenic toxins.

2.2 Introduction

Like many other A-B toxins that mediate their own delivery into cells, high-resolution structures of the enzymatic A-domains (55,96,183) and the receptor-binding portion of the B-domains of glucosylating toxin family members are known (58,184), while the structure and mechanism of the pore-forming translocation domain remains poorly characterized. These inter-connected processes have been proposed to be mediated by the central ~1000 amino acid D-domain (i.e., aa 801-1850), however, with the absence of any structural information for this domain in either the pre-pore or pore state, no framework exists for resolving the functional determinants for this large domain that govern pore formation and translocation. It is well established that in response
to acidic pH, the D-domain undergoes a conformational change that results in the formation of ion-conductive pores in both biological membranes and artificial lipid bilayers (95,97).

It has been hypothesized that the cluster of 172 hydrophobic, highly conserved amino acids in the middle of the translocation domain (i.e., residues 958-1130 in TcdA; and, 956 to 1128 in TcdB) comprised some, if not all, of the segments that form the translocation pore (73). Demonstrating this, however, has been challenging, in large part due to difficulties associated with manipulating Clostridial toxin genes at the genetic level. The recent availability of clones of both TcdA and TcdB in *Bacillus megaterium* expression plasmids, which enable the high-level production of stably folded toxin, has facilitated research in this direction (185,186), however studies specifically addressing the structure and function of the translocation domain have been limited to large-fragment deletions to probe function (108,187).

In the present study, we set out initially with the goal of identifying the determinants of pore-formation and translocation through a comprehensive mutagenesis study using the *B. megaterium* platform. We discovered very early in this pursuit that site-specific mutagenesis of the inherently AT-rich toxin sequence (i.e., G+C = 27%) using the *B. megaterium* system was laborious and inefficient. To address this, we generated a GC-enriched copy of TcdB (i.e., G+C = 45%) with codons optimized for *E.coli* expression, which allowed us to perform high throughput probing of the translocation domain. We identify several single point mutations clustering to within the hydrophobic region of the delivery domain that result in major defects in pore-formation and translocation. We report the unexpected similarity of the identified pore-forming region to that of the translocation domain of DT- and use these data *en bloc* to propose an α-helical model for the translocation pore of TcdB and homologous pathogenic toxins.

### 2.3 Material and Methods

#### 2.3.1 Production of TcdB

#### 2.3.1.1 Expression and purification of recombinant TcdB from *Bacillus megaterium*.

Recombinant TcdB wild type was a *B. megaterium* expression vector pHis1522 encoding the strain VPI10463 obtained from Hangping Feng. Proteins were expressed and purified as previously described (186).
2.3.1.2 Expression and purification of recombinant codon-optimized TcdB constructs.

Codon-optimized TcdB sequence was synthesized (GenScript) to increase GC percentage to 47%. The codon-optimized gene was cloned into an E. coli expression vector pET28a and transformed into E.coli BL21 (DE3) competent cells and expressed as C-terminal His-tagged proteins. 50 mL of overnight culture was inoculated into 1L of LB with 50µg/ml Kanamycin and induced at OD$_{600}$ of 0.6 with 0.5mM IPTG at 37°C for 4 h. Cells were harvested by centrifugation and re-suspended with lysis buffer (20mM Tris pH 8.0, 0.5M NaCl, protease inhibitor) and lysed by an EmulsiFlex C3 microfluidizer (Avestin) at 15,000 psi. After lysing, lysate were centrifuged at 18,000 g for 20 min. Proteins were purified by Ni-affinity chromatography using HisTrap FF column (GE Healthcare). Fractions containing TcdB were verified and pooled with a 100,000 MWCO ultrafiltration device. 10% of glycerol was added; protein concentration was calculated by densitometry (Image Lab 3.0).

2.3.1.3 Mutagenesis of TcdB mutants.

Single point mutations were made in the TcdB codon-optimized sequence using QuikChange lightning multi-mutagenesis kit (Agilent technologies). Sequenced plasmids with confirmed mutations were transformed and expressed using the same conditions as wild type.

2.3.1.4 Small-scale expression of TcdB mutants.

Plasmids expressing TcdB mutants were transformed into E.coli BL21 (DE3) cells. Overnight cultures were prepared in a 24-well block (BD Biosciences) in 5 mL. Cells were harvested by centrifugation and resuspended in buffer (20mM Tris, 500mM NaCl pH 8.0 and protease inhibitor (Sigma) and lysed by lysozyme (Bioshop) as manufacturer’s instructions followed by centrifugation at 4,000 g for 20 min. Supernatant fluids were collected. The concentration of each full-length mutant protein in the lysates was determined by densitometry (Image Lab 3.0).

2.3.2 Characterization Assays

2.3.2.1 Cell viability assay.

CHO-K1 Cells (Chinese hamster ovary cells) were cultured in Ham’s F-12 medium (Wisent) with 10% fetal calf serum (FBS, Wisent) and 1% penicillin and streptomycin (Wisent). CHO-K1 cells were seeded at a concentration of 8,000 cells/well in 96-well CellBind plates (Corning).
The next day, medium was exchanged with serum-free medium and cells were intoxicated by adding TcdB toxins at a serial dilution of 1/3 starting at 1nM. After intoxication, cells were incubated at 37°C, 5% CO₂ for 48 h. Serum (FBS) was added back to cells 24 h after intoxication to a final concentration of 10%. The Cell viability after 48 h was assessed by PrestoBlue® Cell Viability Reagent (Life Technologies). Fluorescence was read on a Spectramax M5 plate reader at an excitation wavelength of 560 nm and an emission wavelength of 590 nm (Molecular Devices).

2.3.2.2 CellTiterGlo ATP assay.

Cell death assay was performed as previously described (164). Briefly, IMR-90 Cells (cultured in EMEM, 10% FBS, 5%CO₂) were seeded in 96-well Cellbind plate at a concentration of 8,000 cells/well. The next day, the growth medium was exchanged with serum free EMEM and incubated at 37°C, 5% CO₂ for 60 min. TcdB toxins were added to cells in dilutions starting at 30 nM. After intoxication, cells were incubated at 37°C, 5% CO₂ for 3 h. The amount of ATP was assessed with CellTiterGlo as per the manufacturer's instructions (Promega). Luminescence was read in Spectramax M5 plate reader at an emission wavelength at 560 nM (Molecular Devices).

2.3.2.3 Rubidium release assay.

86Rb⁺ release assay was performed as previously reported (108) with slight modifications. Briefly, CHO-K1 cells were seeded in 96-well plates in the medium (Ham’s F-12 with 10 % FBS), supplemented with 1 μCi/ml 86Rb⁺ (PerkinElmer) at a density of 1 × 10⁴ cells per well. Cells were incubated at 37 °C, 5% CO₂ overnight. Medium was exchanged with fresh growth medium with 100nM bafilomycin A1 (Sigma) and continued to incubate for another 20 min. Then, cells were chilled on ice and ice-cold medium containing TcdB mutants (10nM) was added. Cells were kept on ice for toxin binding for 1 h at 4°C before they were washed with ice-cold PBS twice to remove unbound toxins. pH-dependent insertion into the plasma membrane was induced by warm, acidified growth medium (37 °C, pH 4.5 or pH 7.5) for 5 min at 37 °C. After 1 hour of further incubation on ice, medium containing released 86Rb⁺ was removed from cell plate and amount of 86Rb⁺ released was determined by liquid scintillation counting with TopCount NXT (PerkinElmer).
2.3.2.4 TNS fluorescence assay.

pH-induced conformational changes of TcdB were assessed as described previously (188). 2 µg of TcdB was prepared in buffer that pH ranging from 4 to 7. 2-(p-toluidiny)-naphthalene-6-sulfonic acid, sodium salt (2,6- TNS, Invitrogen) was added at a final concentration of 150 µM. The final volume was 250 µl and mixed in 96-well black plate (Corning). Mixtures were incubated at 37°C for 20 min. The plate was analyzed in Spectramax M5 plate reader (Molecular Devices) with excitation of 366 nm and an emission scan of 380 to 500nm.

2.3.2.5 Black lipid bilayer experiments.

Lipid bilayer experiments were performed as described previously with modifications (189). Briefly, membranes were made by painting diphytanoyl phosphatidylcholine (Avanti Polar Lipids) in decane across a 200-µm aperture in a Delrin cup by using the brush technique. Both cis and trans compartments contained 1 ml of solutions containing universal bilayer buffer (32) (1 M KCl; 10 mM Tris pH 7.4). Translocation was initiated by adding appropriate amounts of 2 M HCl to the cis compartment to lower the pH to 4.5. Each compartment was stirred continuously throughout the experiment with a small stir bar. Agar salt bridges linked Ag/AgCl electrodes in 3M KCl. The current was amplified through a BC-525C integrating bilayer clamp amplifier (Warner Instruments, Hamden, CT), filtered at a frequency of 0.1 kHz by a low-pass eight-pole Bessel filter and computer-displayed through an analog/digital converter.

2.4 Results

2.4.1 Patterns of Hydrophobicity and Secondary Structure Suggest a Helical Pore for TcdB

To begin to unravel the determinants of pore formation and translocation, we first analyzed the hydrophobicity, sequence conservation and predicted secondary structure elements of the 1050 amino acid translocation domain. Seven stretches of hydrophobicity were identified in TcdB: 985-1005, 1018-1036, 1037-1056, 1064-1089, 1091-1112, 1261-1281, and 1310-1330 (Fig. 2-3A). We excluded the latter two regions since neither were predicted to be hydrophobic in the homologue from Clostridium novyi (TcnA) (Figure 2-1). Notably, the former five hydrophobic segments fell within the "hydrophobic region" of the translocation domain that was predicted previously (i.e., 956-1128). The length of the four hydrophobic segments that all were between
18 to 25 amino acids, combined with the absence of any alternating hydrophobic-hydrophilic "β-barrel" motifs in this region, suggest that the membrane-inserted form of these segments likely adopt an α-helical conformation. When the primary sequence of the hydrophobic region was analyzed using secondary structure propensity algorithms, five α-helical structural elements with four intervening disordered loops were predicted (Figure 2-2).
Figure 2-1 Alignment of the translocation domain of the large clostridial toxin family using ClustalX2.1

Residues 800-1880 are shown using TcdB numbering. The hydrophobic region (956-1128) is highlighted and the regions predicted to be hydrophobic are shown as boxes. Only HH1-HH5 were predicted to be hydrophobic, whereas HH6 and HH7 were predicted in all homologues except for TcnA.
Figure 2-2 Secondary structure prediction for the translocation domain of TcdB using JPRED3.

Numbering is transposed by 800 residues (i.e., 1 is actually residue 801). Predicted helical regions are shown as red rectangles, beta-sheets as green arrows and non-structured areas as lines.
Positing an α-helical mode of membrane insertion, we evaluated the well-characterized α-helical diphtheria toxin (DT) translocation domain for comparison. The hydrophobic helices (TH5-6/7 and TH8-TH9) previously shown to be involved in pore-formation and translocation of DT (190,191) were correctly mapped by this analysis (Fig. 2-3B inset). Unexpectedly, we observed that the general pattern of hydrophobicity was strikingly similar for the 173-residue translocation and the 172-residue hydrophobic region of TcdB. Three peaks of similar length and amplitude were predicted in both toxins. We examine the functional link between DT translocation and the hydrophobic region of TcdB in greater detail below.

![Hydropathy analysis of the translocation domain.](insert-figure)

**(Top)** Hydropathy analysis of the entire translocation domain of TcdB was performed using a membrane protein topology prediction method (TMHMM v2.0) that uses a hidden Markov model to predict transmembrane helices (192). Seven distinct peaks of hydrophobicity are evident; five within the previously described hydrophobic region (73), along with two smaller regions of hydrophobicity between 1280-1350, which were poorly conserved among homologous toxins and thus not pursued. **(Bottom)** Hydropathy analysis of the 172 residue hydrophobic region showing predicted hydrophobic helices (HH1-HH5). **(Inset)** Hydropathy analysis of the 173 amino acid diphtheria toxin translocation domain with established α-helical segments predicted to comprise the translocation pore of DT (181,193).
2.4.2 Validation of a GC-Enriched Toxin B Gene for Mutagenesis and Expression in E. coli

The observation that the putative pore-forming hydrophobic regions of the large translocation domain were localized within the 172 amino acid window led us to investigate which specific amino acids in this region were involved in pore-formation and translocation. To circumvent experimental barriers associated with generating many mutants to the AT-rich Clostridial toxin gene (i.e., G+C = 27%), a copy of the 7,098 base pair TcdB gene was synthesized in which the G+C content was increased to 45%. This mutagenesis-competent copy of TcdB was then cloned into an E. coli expression plasmid in order to enable expression in a host that is more amenable for high throughput characterization than the existing Bacillus megaterium expression system.

To validate the newly constructed GC-enhanced copy of TcdB, we characterized the structure and function of E. coli produced TcdB and compared it to benchmark standards. TcdB produced in E. coli was indistinguishable from TcdB produced in the well-validated B. megaterium system showing equal potency on CHO cells toxicity (Appendix Figure 1). We next measured the cytotoxicity of both purified toxin and soluble clarified lysate from induced E.coli on Chinese Hamster Ovary (CHO) cells, and compared this to uninduced controls, a glucosyltransferase defective mutant (D270A), and WT TcdB produced in B. megaterium (Appendix Figure 1). Purified WT toxins produced in either system yielded similar potency on CHO cells, whereas D270A was similarly inactive when produced in either system. Importantly, WT toxin produced in E.coli clarified soluble lysate was equipotent with the purified toxins (after normalizing toxin concentration of toxin in crude lysates using densitometry). This set of experiments showed that E.coli-produced toxin is functional and further that there are no confounding contaminants in the E.coli preparations as evidenced by the complete lack of toxicity of the uninduced control on CHO cell viability (Appendix Figure 1).

2.4.3 High-Throughput Mapping of the Functional Determinants in the Translocation Domain

With a robust system to probe toxin function in place, we probed residues that were absolutely conserved among the LCT family members, reasoning that functionally important residues would be conserved in homologous toxins. We used a double-mutant strategy in which we mutated each of the residues to both a highly disruptive residue (Lysine), and a more
conservative residue (Cysteine). The highly polar Lysine side chain was selected to increase the probability of identifying a membrane-spanning segment; introducing the polar and charged Lys side chain into a marginally hydrophobic membrane-inserting segment could be expected to prevent insertion of this segment and thus pore-formation. On the other hand, Cysteine, like Alanine, is a relatively benign substitution that can both help identify key functional residues and has the added downstream benefit of offering the possibility of attaching sulfhydryl probes in TcdB for structure/function studies.

Figure 2-4. High throughput mapping of the functional determinants in the translocation domain of TcdB.

Functional consequences of Cys and Lys substitutions in the hydrophobic region of TcdB. Mutant soluble lysates were titrated onto CHO cells (using 3-fold dilutions) in 96-well plates and incubated for 48h at 37°C (n=4). In parallel, an aliquot of each mutant was used to measure the concentration using band densitometry after SDS-PAGE. 48h later cell viability of treated cells was quantitated by measuring PrestoBlue® fluorescence using a SpectraMax M2 fluorescence microplate reader. (inset) Sample titration curves of WT TcdB and L1041K mutant TcdB. Grey shading represents the wildtype-like range of activity (i.e., ± 5-fold wildtype TcdB).
The impact of each mutation on TcdB function was quantified by measuring the dose-dependent reduction in cell viability 48 hours post-toxin addition relative to wild type TcdB (Fig. 2-4). Of the nearly 90 mutants generated, only Y971 and L1048 were not expressed upon mutation, and thus were not evaluated further. Of the remaining mutants, we identified eight residues that when mutated, gave rise to a greater than a 90% reduction in toxicity relative to WT (Fig. 2-4). As expected, for virtually all sites tested, mutation to Lysine was more detrimental to function than mutation to Cysteine. Four positions (i.e., D1037, G1098, I1099, and L1106) displayed a greater than a 99% reduction in toxicity of which L1106K was the most defective with an observed shift of over 1000-fold relative to wild type TcdB. Mutants displaying greater than a 10-fold reduction in toxicity were re-expressed in large scale and purified using Ni⁺-affinity chromatography and/or anion exchange chromatography and re-tested in triplicate. We observed excellent overall correlation in potency between the soluble lysates and purified toxins, which both confirmed our screening results and further validated our soluble lysate screening approach.

2.4.4 Pinpointing the Nature of the Defects in TcdB that Diminish Function

Defective TcdB mutants were studied in detail to determine the mechanistic basis for their defects. To rule out the possibility that point mutations were causing gross and global misfolding of toxins - thus rendering them unable to intoxicate cells - we evaluated the state of folding for each mutant using the hydrophobic dye TNS, which displays increased fluorescence when binding to hydrophobic patches of polypeptide (i.e., unfolded proteins). In addition to confirming that all mutants tested were folded, these studies show that the pH dependence of unfolding was preserved for mutant toxins (Appendix Figure 2). To address this further, we evaluated the enzymatic activity of the glucosyltransferase domain for activity in each defective mutant. All mutants tested showed comparable activity to WT TcdB with specific activities that were within ± 2-fold wildtype levels (Appendix Fig. 2), further suggesting that mutant toxins were otherwise folded.
Figure 2-5. Characterization of defective purified TcdB mutants

(A) Pore formation on biological membranes. Pore formation of purified mutant toxins was tested on CHO cells pre-loaded with $^{86}$Rb$. Pore formation was induced by acidification of the external medium (control pH 7.0; black bars, pH 4.5; grey bars) - see Methods for details of assay (n=5). (B) Pore formation on planar lipid bilayers. ~100 pM of WT or L1106K toxin was added to the cis chamber of planar DiPhPC/n-decane membranes in buffer containing 1M KCl. After 2 minutes the pH of the cis chamber was dropped to pH 4.5 using a defined concentration of dilute HCl. Measurements were performed with $\Psi$ = +50mV, cis-positive) at room temperature. (C) Effect of pore formation on enzyme-independent cytotoxicity. The high-dose acute cytotoxicity of purified WT TcdB, a glucosyltransferase-defective mutant (D270A), and a pore formation defective mutant (L1106K). Constructs were tested on human IMR-90 fibroblasts under necrosis-like conditions as described previously (164).
Defective mutants were then tested for translocation domain-specific functions. We first tested the ability of each of the defective mutants and control toxins to release $^{86}$rubidium ions from CHO-K1 cells upon binding to the cell surface and acidification of the medium to trigger insertion into the plasma membrane (Fig. 2-5A). Wild type TcdB from all sources was able to form pores and release rubidium upon acidification to levels comparable to the detergent controls. As expected, the glucosyltransferase inactive mutant D270A, and a construct in which the entire GTD was removed (i.e., ΔGTD) were also able to form pores at low pH. In contrast, all of the defective mutant toxins tested showed defects in pore-formation, with most showing levels comparable to the no toxin control. Interestingly, two mutant toxins, 1037K and 1038K showed intermediate levels of pore-formation, suggesting that pore-formation was reduced, but not ablated under these conditions. Similarly, the previously identified mutant E970K/E976K, which to date is the only other mutant reported to date that affects pore-formation, showed a partial, but not complete reduction in pore-formation. In parallel, we tested pore-formation in synthetic lipid bilayers using electrophysiological methods. As shown in Fig. 2-5B (top trace), WT TcdB induced an increase in membrane activity at acidic pH (pH 4.5) with a holding potential of +50mV ($\Delta \psi = +50$ mV, cis-positive). The observed large and transient currents were reproducible and similar in character to previous planar lipid bilayer experiments (108). By contrast, despite several attempts ($n = 8$), the use of higher concentrations and extended traces, we were unable to detect any channel activity for the L1106K mutant consistent with an inability to form ion-conductive pores (Fig. 2-5B).

2.4.5 Examining the Role of Pore-Formation on TcdB-Induced Necrosis

The experimental conditions that we used in this study to measure TcdB cytotoxicity depend on functional autoprocessing and glucosyltransferase domain functions. Recently, Lacy and coworkers identified a second, alternative mode of cytotoxicity for TcdB that was found to be independent of autoprocessing and glucosyltransferase functions (164). At higher doses of TcdB – which are hypothesized to be possible during infection – cells undergo a rapid necrotic-like cell death via an NADPH oxidase pathway, characterized by a rapid depletion of ATP (194). We took advantage of the pore-defective mutants identified here to examine the importance of pore-formation on this alternative mode of TcdB-mediated cell death. Whereas both WT TcdB, and the glucosyltransferase inactive D270A mutant equally induced a rapid depletion of ATP at 100 pM, we saw no reduction in cellular ATP for L1106K up to 100 nM, indicating that pore-
formation is indeed required for the TcdB-mediated necrosis (Fig. 2-5C). In support of this, Donald et al., recently showed using the milder double mutant E970K/E976K, which partially reduces pore-formation, shifted the potency of a TcdB variant with attenuated GTD and CPD activity by ~100-fold. The difference in impact is consistent with our finding that under equivalent conditions, L1106K has a greater impact on pore formation than E970K/E976K.

2.4.6 Mapping the Determinants of Pore formation onto a Model of the TcdB Pore

To place these functional data onto a structural framework, we considered a model of the membrane-inserted form of TcdB based on the predicted hydrophobicity and secondary structure of the hydrophobic region and using knowledge of DT to help orient the model. Mapping the residues that we identified as being sensitive to mutagenesis onto the TcdB model revealed two "hotspots" that conspicuously localized to the distal side of the membrane (Fig. 2-6).

![Figure 2-6. Proposed model for TcdB translocation.](image)

Mapping the functional determinants of pore-formation and translocation onto a working model of the TcdB translocation pore and the model of the "double-dagger" DT pore proposed originally by Choe et al., (181) and later supported by Wang and London (193).

In forming the pore, these residues would be expected to traverse the greatest distance into the membrane relative to the proximal side of the membrane where the pre-pore residues before insertion. Intriguingly, when we mapped the four residues in the DT translocation domain previously shown to reduce DT-mediated toxicity by more than 100-fold, we observed a similar phenomenon; defective mutants localize to the distal end of the pore. In support of this model, we found shared residues within the hotspots that appear to be functionally important in both
toxins. Asp-1037 in TcdB and Asp-295 in DT, though offset slightly in primary sequence alignments are positioned in the loop between the first two membrane-spanning helices and mutations to Lysine in both cases as shown here for TcdB and previously with DT (195) resulting in a >100-fold shifted relative to WT toxin. In the loop region intervening the second helical hairpin hotspot, Pro-345 in TcdB is aligned with Pro-1095 in DT. Pro-345 in DT was previously shown to prevent pore formation and membrane insertion resulting in a 99% reduction in toxicity to Vero cells (196). Our mutagenesis studies show that the P1095K mutation also prevents pore formation similarly giving rise to a ~95% reduction in toxicity (Fig. 2-4). Finally, at the heart of the membrane-insertion region are the well-studied pore-formation/translocation defective DT mutants E349K and D352K, which were shown previously to reduce functional toxicity in Vero cells by >100-fold due to an impairment in membrane-insertion of the TH8-TH9 helical hairpin (195,197). Sandwiched between these two residues are L350 and V351 in DT, corresponding to the two defective mutants, L1106 and V1107 uncovered here in TcdB.

Four hydrophobic segments of ~20 amino acids are proposed to span the lipid bilayer in the pore state as two helical hairpins - similar to DT. Functionally important residues affecting membrane-insertion/pore-formation map to the cytosolic face of the membrane in TcdB and DT. (Below) A ClustalX alignment of TcdB, TcdA and DT shows regions of conservation in the two pore-formation "hotspots" with defective mutants indicated as arrowheads (for TcdB) or diamonds (for DT). Residues within the boxed region (i.e., S1105-N1108 for TcdB, S1107-N1110 for TcdA and E349-D352 for DT) are highly susceptible to mutagenesis.

2.5 Discussion

Research over the past decade has provided tremendous insight into the structure and function of the pathogenic toxins of *C. difficile*. Despite these significant advances, there remains a large gap in our understanding of the underlying structural and functional features that explain how the central translocation domain mediates the critical step of delivering the cytotoxic glucosyltransferase domains across the endosomal membrane into the cytosol. In this study we provide the first comprehensive analysis of the functional determinants of the translocation domain and offer a model of the TcdB pore. Using a synthetic GC-enriched copy of TcdB, we identified eight residues, between residues 1035 and 1107 that when mutated resulted in a greater than 90% decrease in function. One of these pore-formation defective mutants, L1106K, reduced
TcdB toxicity by greater than 1000-fold relative to WT toxin. These key functional determinants are conserved in the members of the LCT family as well as among the different variants of TcdB (Appendix Fig 3). That these studies were conducted under conditions where autoprocessing and glucosyltransferase functions are required for cytotoxicity, argues that formation of this pore is required for translocation of the glucosyltransferase effector into cells.

Recently, Genisyuerek et al. generated a series of internal-, amino-, and carboxy-terminal truncations of the translocation domain in order to delineate the determinants of pore formation and translocation for TcdB (108). Consistent with our findings and our model, they concluded that residues 1-1500 (with a receptor-binding domain) encompassed the translocation machinery enabling functional toxicity. Importantly, however, they also showed that a small fragment (i.e., 830-1025) was able to form channels in planar lipid bilayers (108). Since this fragment only partially overlaps with the residues in our model, and excludes the most hydrophobic helices within the delivery domain, this finding would appear to contradict our model; however, it is important to note, as the authors of this study also suggest, that the appearance of ion channel activity in a synthetic lipid bilayer with fragments of a toxin, does not necessarily indicate the existence of the minimal translocation domain. Distinctions between the minimal channel forming region and the minimal translocation domain have been made previously for DT. A C-terminal fragment of DT, containing only half of TH8 and all of TH9 formed channels in lipid bilayers that had similar electrophysiological properties as those formed by wild-type toxin (198), despite lacking the necessary determinants for translocation.

Taken together, the data presented here support a model for the TcdB pore that is similar in principle to the double-hairpin model has been suggested previously for DT (181,193). As with DT, we found that introducing a positive charge in regions that must travel furthest across the apolar bilayer during insertion to be the most detrimental to function. Whether the four transmembrane segments proposed require further oligomerization for functional translocation is not known at present. Furthermore, how the hydrophobic helices are concealed within the large 1050 amino acid translocation domain pre-pore structure prior to insertion is unknown and awaits a high-resolution structure of this elusive domain.

Our observation that single point mutants within the translocation domain of TcdB could prevent pore formation and translocation additionally highlights the importance of this domain in
pathogenesis. Several recent studies have shown that under certain conditions, the glucosyltransferase domain (164), the auto protease domain (114,164) and even the receptor-binding domain (66) are dispensable for function. Our panel of pore-defective mutants identified here, uniquely allowed us to test the importance of translocation domain function in toxin function. We found that pore-formation defective mutants are non-toxic under all toxic paradigms tested, including cytopathic function and cytotoxic function both at low (GTD-dependent) and high concentrations (GTD-independent) (164). Along with providing a mechanistic basis for understanding how TcdB may elicit cytotoxicity in the absence of enzyme activity, these findings also offer a new avenue for therapeutic intervention to halt the progress of these pathogenic toxins with small molecules or monoclonal antibodies. Agents that prevent the pre-pore to pore conversion, or block proper insertion/assembly of the pore can be expected, based on our findings, to fully inhibit toxin pathogenesis. Furthermore, given the magnitude of the reduction in toxicity combined with the subtly of a single point mutation on holotoxin folding (as demonstrated), the defective single-point mutants described herein offer ideal candidates for vaccine and/or monoclonal antibody development to neutralize these potent toxins. As these mutants have thus far only been tested in vitro on cultured mammalian cells, confirmation of the exciting potential therapeutic applications of these mutants awaits future in vivo validation studies in a disease-relevant model of infection.
Chapter 3

Defective Mutations within the Translocation Domain of
Clostridium difficile Toxin B Impair Disease

The work presented in this chapter was published in *Pathogens and disease* journal.


**Contributions:**

My contributions to this work included designing and making of the expression constructs, protein purification and designing and performing cell biology and biochemical studies for characterization. The *in vivo* mouse study was carried out by T. Hamza, Feng lab (Figure 3-1B). R. Melnyk supervised the project.
3 Defective Mutations within the Translocation Domain of *Clostridium difficile* Toxin B Impair Disease

### 3.1 Overview

The *Clostridium difficile* toxins A and B are the main virulence factors for symptoms of *C. difficile* disease and targets for vaccine development. Several studies showed the importance of toxin B alone in the pathogenesis of *C. difficile* infection. High-throughput mapping of the functional determinants in the translocation domain of *Escherichia coli* codon-optimized TcdB recently revealed crucial residues in pore formation and toxin translocation. In this study, we investigated the effects of mutating a critical site involved in pore-formation, Leu-1106, to residues different size and polarity (Phe, Ala, Cys, Asp). We observed a broad range of effects on TcdB function *in vitro* consistent with the role of this site in pore-formation and translocation. We show that mice challenged systemically with a lethal dose (LD100) of the most defective mutant (L1106K) showed no symptoms of disease highlighting the importance of this residue and the translocation domain in disease pathogenesis. These findings offer perspectives for translocation pore structure and pharmacological strategies against *C. difficile* infections.

### 3.2 Introduction

Our previous work from chapter two focused on identifying the elements in pore formation of TcdB. We have identified several single point mutations within the hydrophobic region of TD that result in major defects in pore formation and translocation. Among those point mutations, position L1106K displayed the most defective in toxicity with over 3 logs relative to wild-type TcdB (107). L1106K mutant showed a defect of pore-formation both in an *in vitro* system (planar lipid bilayer electrophysiology), and on biological membranes (*86* rubidium release of ions from CHO-K1 cells upon binding to the cell membrane). In this study, we further investigated the role of defective pore-formation mutations within translocation of TcdB in CDI pathogenesis.

### 3.3 Material and methods

#### 3.3.1 Mouse systemic toxin challenge

Six- to eight-week-old female CD1 mice were purchased from Harlan Laboratory (IN, USA). Mice were housed in pathogen-free facilities. The animal study was carried out in strict
accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the Committee on the Ethics of Animal Experiments of the University of Maryland Dental School (protocol #06-14-007). Mice were euthanized when moribund after toxin challenge, and all efforts were made to minimize suffering. To assess in vivo toxicity of the toxins (TcdB and L1106K), groups of mice were challenged intraperitoneally (IP) with either lethal dose (LD\textsubscript{100}) of wild-type TcdB (5µg/kg mouse) or (0.5g/kg mouse) of L1106K (100x LD\textsubscript{100} of TcdB). Mouse survival was monitored and data were analyzed by Kaplan–Meier survival analysis with Log rank test of significance.

### 3.3.2 Differential scanning fluorometry

Differential scanning fluorometry was performed in a similar manner as described previously (199). TcdB protein was diluted in phosphate buffer (100 mM KPO4, 150 mM NaCl, pH 7) containing 5x SYPRO Orange (Invitrogen). A Bio-Rad CFX96 qRT-PCR thermocycler was used to establish a temperature gradient from 15°C to 95°C in 30s increments, while simultaneously recording the increase in SYPRO Orange fluorescence as a consequence of binding to hydrophobic regions exposed on unfolded proteins. The Bio-Rad CFX Manager 3.1 software was used to integrate the fluorescence curves to calculate the melting point.

### 3.3.3 Glucosyltransferase activity of the toxins

The wild type and mutant toxins were assessed for glucosyltransferase activity by assaying glucosylation of the Rho GTPase Rac1, both intracellularly and in cell-free assays. In intracellular assay, Vero cells in 12-well plates were exposed to different concentrations of toxins for 4 h at 37 °C. Cells were lysed with SDS sampling buffer. In cell-free assays, Vero cell pellets were re-suspended in glucosylation buffer (50 mM HEPES, pH 7.5, 100 mM KCl, 1 mM MnCl\textsubscript{2} and 2 mM MgCl\textsubscript{2}) and lysed with a syringe (25G, 40 passes through the needle). After centrifugation (15,000 rpm, 15 min), the cell lysate was incubated with different doses of TcdB, or L1106K at 37 °C for 1 h. The reaction was terminated by heating at 100 °C for 5 min in SDS-sample buffer. In both cell-free and intracellular assays, Rac1 glucosylation was detected using antibody that specifically recognizes the non-glucosylated form of Rac1 (clone 102, BD Bioscience). Anti-β-actin (clone AC-40, Sigma) was used to detect β-actin and ensure that samples were loaded evenly on the SDS polyacrylamide gels.
3.4 Results

3.4.1 Generation of the functional residues at amino acid 1106

CHO-K1 cells were treated with TcdB wild-type or mutants for 48 h and cell viability was assessed by measuring the fluorescence of cells treated with PrestoBlue cell viability reagent. Substitutions to hydrophobic residues of any size (Leu, Ala, Phe) had the least impact on TcdB translocation, followed by polar/uncharged (Cys), followed by acidic (Asp), followed by basic substitutions (Lys) (Figure 3-1A). This is consistent with the idea that this residue is a key site for membrane insertion; the more polar the residue at this position, the greater the impact on pore formation/translocation and thus intoxication.

Figure 3-1. *In vitro* and *in vivo* toxicity of mutant TcdB.

(A) CHO-K1 cells were treated with TcdB wild-type or mutants and cell viability was measured by fluorescence of cells treated with cell viability reagent (PrestoBlue). (B) CD1 mice were challenged IP with either LD$_{100}$ of toxins (5 µg/kg mouse) or 100× LD$_{100}$ of L1106K (0.5
g/kg mouse). Mouse survival was monitored and data were analyzed by Kaplan–Meier survival curve with Log-rank test of significance. \( n=5, P<0.0001 \), between wild-type and mutant toxin groups).

3.4.2 L1106K mutation diminishes its activity in vivo

Reduced cellular in vitro toxicity of L1106K mutant led us to assess toxicity of wild-type TcdB and the most defective mutant-L1106K by challenging the mice systemically (IP). Mice challenged with LD\(_{100}\) (5\( \mu \)g/kg mouse) of TcdB developed signs of systemic disease rapidly and all became moribund and died within six hours. In contrast, none of mice challenged with L1106K mutant at the same dose showed any systemic disease symptoms, while mice challenged with 100x LD\(_{100}\) (0.5g/kg mouse) of L1106K mutant showed signs of disease 48 h post challenge (Figure 3-1B). These data clearly demonstrate that mutation at 1106 residue is critical for TcdB toxicity and L1106K mutant is substantially attenuated in in vivo toxicity.

3.4.3 Characterization of L1106X activities in vitro

In attempt to ascribe a molecular mechanism to functional consequences observed on toxin function, we conducted further in vitro studies on the various L1106X mutants. Wild type toxin and mutant were tested for the ability to release \(^{86}\)Rubidium ions from CHO-K1 cells upon binding to the cell surface and acidification of the medium to trigger insertion into the plasma membrane. As shown in Figure 3-2A, a remarkable range of effects on pore formation are seen upon substituting residues position 1106. To show that these effects were not due to toxin misfolding, the protein thermal stability of each toxin was tested (Figure 3-2B). For the translocation ability of TcdB and L1106K mutant, we assayed Rac1 glucosylation intracellularly and in cell-free assay. Glucosylated Rac1 levels in Vero cells did not change with increasing concentration of TcdB mutant (Figure 3-2C). On the other hand, L1106K showed similar activity in Rac1 glucosylation in cell-free assay (Figure 3-2D).
Figure 3-2. Characterization of mutant TcdB.

(A) Pore formation on biological membranes. Pore formation of purified mutant toxins was tested on CHO cells preloaded with $^{86}\text{Rb}^+$. Pore formation was induced by acidification of the external medium (control pH 7.5; black bars, pH 4.5; gray bars)—see the section ‘Materials and Methods’ for details of assay ($n = 2$). (B) Temperature-dependent fluorescence measurements of TcdB melting temperature (Tm). Values represent the mean ± SD from three independent experiments. (C) Glucosyltransferase activity of the mutant toxins. Vero cells were exposed to different concentrations of TcdB-WT or L1106K mutant for 4 h. (D) Vero cell cytosolic fraction was collected and exposed to wild-type and mutant toxins at the indicated concentrations for 1 h. In both C and D, western blot was performed using monoclonal antibody (Clone 102) that only binds to non-glucosylated Rac1. β-actin was used as an equal loading control.
3.5 Discussion

TcdB is one of the essential virulence factors of *C. difficile* as a number of pathogenic, clinically-isolated strains express functional TcdB alone (49,200). The role of enzymatic domains of TcdB in disease pathogenesis has been studied recently (83,201); however, information regarding the translocation domain and pore formation in pathogenesis has not yet been reported.

In this study, we demonstrate that defective variants of TcdB are defined by impaired pore formation, which results in reducing toxicity to cells and animals. We have identified several residues located between amino acids 1,035 and 1,107 that when individually mutated, markedly reduced the toxic activity by over 1000-fold. Therefore, this sensitive segment appears to be a crucial structure for the function of TcdB translocation and pore formation (107). The observation that L1106K toxicity is substantially reduced *in vitro* and *in vivo*, yet maintains its cellular binding, internalization, and glucosyltransferase activities (107) highlights the importance of this domain in translocating TcdB across the endosomal membrane into the cytosol and CDI pathogenesis.

We have previously demonstrated in Chapter 2 that pore-defective mutants have reduced toxicity in cell culture (107). In this study we further assessed the role of residues differing in size and/or polarity at 1106 on function. Reduced cellular toxicity was seen in cells treated with hydrophobic residues (Leu, Ala, Phe), followed by Cys, then Asp, followed by basic substitutions (Lys) (Figure 2-1A). This is consistent with the idea that this residue is a key site for membrane insertion; more polar, more defective in pore formation resulting in an inability to intoxicate cells.

*C. difficile* infection is the leading cause of healthcare-associated diarrhea in Europe and North America (202, 203). Colonization of *C. difficile* the intestinal mucosa leads to toxins production and leakage to the systemic circulation (204, 205) causing adverse effects including causing cytoskeletal disruption, diarrhea, colitis, and death (141). In this study, we observed typical signs of systemic disease in mice challenged with LD$_{100}$ (5µg/kg mouse) of wild-type TcdB while a similar dose of L1106K exhibited no disease symptoms. Only at a dose of 100x LD$_{100}$ (0.5g/kg mouse), L1106K caused 40% death rate. Thus, mutations in TD not only hold a great promise to understand the mechanism of toxin translocation through pore formation, but also provide
insights in the development of putative vaccines against hypervirulent multi-drug resistant bacteria such as *C. difficile*. 
Chapter 4

Crystal Structure of *Clostridium difficile* Toxin A

The work presented in this chapter was published in *Nature Microbiology* journal.


**Contributions:**

My contributions to this work included characterization of mutant proteins by cell viability and rubidium release assays (Figures 4-3e, 4-3f) in collaboration with Lacy lab. The Lacy lab (N.M. Chumbler, S. A. Rutherford and D. B. Lacy) crystallized, determined the structure of TcdA1-1832 and performed most of the biochemical experiments and data analysis. J.P. Lisher performed ICP-MS assay. E. Farquhar performed XAS measurement.
4 Crystal Structure of \textit{Clostridium difficile} Toxin A

4.1 Overview

\textit{Clostridium difficile} infection is the leading cause of hospital-acquired diarrhea and pseudomembranous colitis. Disease is mediated by the actions of two toxins, TcdA and TcdB, which cause the diarrhea, as well as inflammation and necrosis within the colon (4,206). The toxins are large (308 and 270 kDa, respectively), homologous (47% amino acid identity) glucosyltransferases that target small GTPases within the host (141, 207). The multidomain toxins enter cells by receptor-mediated endocytosis and, upon exposure to the low pH of the endosome, insert into and deliver two enzymatic domains across the membrane. Eukaryotic inositol-hexakisphosphate (InsP6) binds an autoprocessing domain to activate a proteolysis event that releases the N-terminal glucosyltransferase domain into the cytosol. Here, we report the crystal structure of a 1,832-amino-acid fragment of TcdA (TcdA\textsubscript{1832}), which reveals a requirement for zinc in the mechanism of toxin autoprocessing and an extended delivery domain that serves as a scaffold for the hydrophobic \(\alpha\)-helices involved in pH-dependent pore formation. A surface loop of the delivery domain whose sequence is strictly conserved among all large clostridial toxins is shown to be functionally important, and is highlighted for future efforts in the development of vaccines and novel therapeutics.

4.2 Material and Methods

4.2.1 Protein expression and purification

GST-Rac1 was expressed and purified as described previously (96). Toxin expression plasmids were transformed into \textit{B. megaterium} protoplasts according to the manufacturer's protocol (MoBiTec). Transformants were grown in Luria-Bertani (LB) broth containing 10\(\mu\)g ml\(^{-1}\) \(\mathrm{tetracycline}\) at 37\(^\circ\)C and 220 r.p.m. overnight to produce a seed culture. To 1L of LB, 30ml of the overnight seed was used as inoculum. The inoculated cultures were grown at 37\(^\circ\)C until an optical density of 0.3–0.4 at 600 nm was reached. Protein expression was induced using 5g l\(^{-1}\) of D-xylose solid (TCI, X0019). After a further \(~4\)h more at 37\(^\circ\)C and 220 r.p.m., the cells were harvested into 1L bottles at 4\(^\circ\)C and 5,000g for 30min. Pellets were resuspended in buffer containing 20 mM Tris, pH8.0, 300mM NaCl, 10\(\mu\)g ml\(^{-1}\) DNaseI, protease cocktail (Sigma, P8849) and 20\(\mu\)g ml\(^{-1}\) lysozyme. The suspensions were homogenized using a dounce
homogenizer and then lysed at room temperature at 25,000 psi (Constant Cell Disruption Systems). The lysates were placed on ice then centrifuged at 18,000 r.p.m. in a JA-20 fixed-angle rotor for 25 min at 4°C. After filtering the chilled supernatants through 0.22 mm filters, the proteins were purified using nickel affinity chromatography at 4°C. Further purification was performed at room temperature using anion exchange chromatography followed by gel-filtration chromatography into either 20 mM Tris, pH 8.0, 100 mM NaCl (for crystallization) or 20 mM HEPES pH 6.9, 50 mM NaCl for cell-based experiments.

4.2.2 Crystallization

TcdA<sub>1832</sub> and S1329C TcdA<sub>1832</sub> were concentrated to 10 mg ml<sup>−1</sup> in 20 mM Tris, pH 8.0, 100 mM NaCl. Crystallization was performed using the hanging drop method at 21 °C with a 1:1 ratio of protein to mother liquor. The mother liquor formulation for wild-type (WT) crystals was 100 mM Bis-Tris, pH 6, 11% PEG 4000, 30–50 mM guanidium chloride (GuCl). The mother liquor formulation for the S1329C crystals was 100 mM Bis-Tris, pH 5.8, 8% PEG 4000, 50 mM GuCl. Crystals were exchanged into appropriate mother liquor containing 20% glycerol, mounted on cryo loops and flash-cooled in liquid nitrogen.

Heavy atom derivatives of TcdA<sub>1832</sub> were prepared by soaking crystals in the appropriate mother liquor containing either 5 mM mercuric chloride for 90 min, 5 mM mercuric chloride for 3 days, 1 mM gold (III) chloride hydrate for 40 min, or 1 mM K<sub>2</sub>PtCl<sub>2</sub> for 40 min. Heavy atom derivatives of S1329C TcdA<sub>1832</sub> were prepared by soaking crystals in 5 mM mercuric chloride for 3 days.

4.2.3 Structure determination and refinement

X-ray data were collected from single crystals on LS-CAT beamline 21 ID-D at the Advanced Photon Source (Argonne, IL) at 100K. Diffraction data were indexed, integrated and scaled using X-ray Detector Software (XDS) (208) or HKL2000 ((209) Appendix Table 1). The two mercury data sets were compared to the native data set using multiple isomorphous replacement with anomalous scattering in SHARP (210). The analysis revealed five mercury sites in the two mercury data sets, differing only in their occupancies, and was consistent with the expectation that each protein monomer would have five free cysteine residues. The heavy atom positions were used to calculate initial phases, which were included in an auto-building protocol in
PHENIX (211). The fragments generated by auto-building guided manual placement of the apo-GTD structure (PDB ID 3SS1) (96). Phases from the GTD model were combined with the phases from SHARP to calculate a new map and initiate a new round of autobuilding. The fragments generated through autobuilding allowed for manual placement of the APD (PDB ID 3H06) (57). Phases from the combined GTD and APD model were combined with the phases from SHARP to calculate a new map and initiate new rounds of automated and manual building. Further phase improvement came from multi-crystal averaging. The working model (consisting of the GTD, most of the APD and a series of unconnected fragments from the delivery domain) was used as search model for molecular replacement into the native, platinum and gold data sets. The models and phases from each data set were subjected to multi-crystal averaging and density modification in PHENIX and resulted in excellent quality maps. One area of ambiguity was resolved through site-specific introduction of a mercury atom: crystals of a S1329C TcdA\textsubscript{1-1832} mutant were derivatized with mercuric chloride, and the sixth heavy atom site was identified using PHENIX. The model was generated through an iterative process of manual building in Coot (212) and refinement using Phenix (211). The final model reflects the 50–3.25Å native data set ($R$-factor = 18.2%, free $R$-factor = 23.7%) with 92.3% of the residues in the most favoured regions of the Ramachandran plot with 0.6% outliers. The model contains residues 4–944 and 951–1802 along with one zinc atom.

4.2.4 X-ray absorption spectroscopy

XAS experiments were carried out at beamline X3B of the National Synchrotron Light Source, which was equipped with a sagitally focused Si (111) double-crystal monochromator and a nickel-coated mirror for harmonic rejection. A helium Displex cryostat was used for temperature control ($\sim$15 K typical sample temperatures). Fluorescence detection was provided by a 31-element solid-state germanium detector array (Canberra). Samples of TcdA (10 mg ml\textsuperscript{-1}) and buffer blanks were loaded into 30 µl polycarbonate cuvettes wrapped in 1mil Kapton tape and then frozen by immediate immersion in liquid nitrogen. The K\textalpha fluorescence emission spectra from TcdA and buffer samples in the X-ray beam (incident energy = 10 keV) were examined. There was a significant increase in the total zinc fluorescence counts for the TcdA sample compared with buffer, while fluorescence for the Mn–Cu series was unchanged. XAS measurements were therefore carried out at the zinc K-edge on TcdA, over an energy range of 9.46–10.3 keV. Internal energy calibration was provided by simultaneous measurement of a zinc
metal foil, with the first inflection point of the edge set to reference energy of 9,659 eV. Calibration and averaging of XAS data were carried out using Athena (213)

4.2.5 ICP-MS

Proteins were prepared as described above and dialysed overnight into metal-free buffers: 20 mM HEPES pH 6.9, 50 mM NaCl or 100 mM Bis-Tris pH 6.0, 50 mM NaCl at 4 °C. Samples analysed in the presence of InsP6 were dialysed in buffer containing 10 mM InsP6. Samples analysed in the presence of TPEN were dialysed in 1 mM TPEN for 8 h at room temperature. Buffers containing 10 µM ZnCl$_2$ and 1 mM TCEP were used to add zinc back to the protein. Protein samples were analysed for metal content by using 50µl of the protein solution and diluting in 2.5% (vol/vol) nitric acid (Sigma-Aldrich, TraceSELECT quality) to a final volume of 3ml for ICP-MS analysis. In samples with significant precipitation after acidification, the samples were centrifuged at 15,000 g for 20 min to pellet any precipitate, and the solution was transferred to a fresh tube for measurement. The diluted samples were analyzed for $^{66}$Zn, $^{55}$Mn, $^{63}$Cu and $^{60}$Ni using a 1–30 ppb standard curve using stock solutions (Perkin Elmer). Analyses were performed using a PerkinElmer ELAN DRCII ICP-MS. The instrument was equipped with a Microflow PFA-ST concentric nebulizer with a 100 µl min$^{-1}$self-aspiration capillary, a cyclonic spray chamber, a quartz torch and nickel sampler/skimmer cones. Germanium at 50 ppb was added as an internal standard using an EzyFit glass mixing chamber. Concentrations (in ppb) were corrected for the dilution factor, and the molar concentrations and molar ratios ($^{66}$Zn/protein) were determined for each sample.

4.2.6 Autoprocessing assays

Assays were performed as described previously (164). Reactions testing the effect of TPEN were pre-incubated with 10mM TPEN at 37°C for 2h before the addition of InsP6. Autoprocessing assays with samples also analyzed by ICP-MS were performed in the appropriate dialysate and indicated reactions contained 100nM InsP6 pre-treated with TPEN to remove zinc from the InsP6 stock solution. Gels were quantified using ImageJ (214). The intensity associated with cleaved GTD was divided by the intensity for the intact protein in the absence of InsP6.
4.2.7 Cell lines

Chinese hamster ovary (CHO-K1), HeLa and Vero cells were authenticated and verified to be mycoplasma-free at the time of purchase from ATCC and were aliquoted into primary frozen stocks. Experiments were conducted in cells passaged less than 30 times from the frozen stock.

4.2.8 Cell binding and Rac1 glucosylation in cells

HeLa cells were synchronized by cooling to 4°C and then intoxicated with 10 nM toxin or buffer. The cells were returned to 4°C for 1h and then shifted to 37°C for 3h. The cells were harvested and lysed (250 mM sucrose, 10 mM Tris pH 7.5, 3 mM imidazole), samples were boiled, and proteins were separated by SDS–polyacrylamide gel electrophoresis (PAGE). Samples were analyzed by western blot with primary antibodies specific for TcdA CROP (Abcam, ab19953), unglucosylated Rac1 (BD, 610650) and total Rac1 (Millipore, clone 23A8). Binding of an anti-mouse, HRP-conjugated secondary antibody (Jackson ImmunoResearch Laboratories, 115-035-174) was detected with a LumiGLO kit (Cell Signaling) according to the manufacturer's instructions.

4.2.9 Statistical analyses

No statistical method was used to predetermine the sample size. Error bars displayed throughout the manuscript represent standard deviation (s.d.) and were calculated from biological replicates. No data were excluded from analysis. The number of replicates for each experiment is indicated in the figure legends.

4.3 Results and Discussion

Although efforts to obtain well-diffracting crystals of either the TcdA or TcdB holotoxins have been unsuccessful, low-resolution structures of TcdA and TcdB have been determined by electron microscopy (EM) and small-angle X-ray scattering, respectively (56, 215). EM analysis of TcdA and TcdB revealed that the elongated solenoid structure of the CROP (58,84,184) can adopt multiple conformational states relative to the rest of the protein (56), so we generated a TcdA construct with the CROP deleted (TcdA1832) for crystallization. The structure was determined and refined to 3.25 Å resolution (Appendix Table 1 and Appendix Fig. 4), and reveals significant interactions between the GTD and APD and an extended and topologically complex delivery domain (Fig. 4-1b). The last residue visible in the TcdA1832 structure is S1802.
Clear placement of the TcdA<sub>1-1802</sub> structure into the holotoxin EM structure (56) (Fig. 4-1d) indicates that the CROP extends from the base of the APD and could impact InsP6 binding. This is consistent with reports that the CROP interacts with N-terminal sequences of TcdA to repress autoprocessing activity (118, 216) and our observation that a shorter construct (TcdA<sub>1795</sub>) undergoes autoprocessing more efficiently than TcdA and TcdA<sub>1832</sub> (Appendix Fig. 5).
**Figure 4-1. Structure of TcdA.**

a, The TcdA primary structure can be divided into four functional domains: the glucosyltransferase domain (GTD, red), the autoprotease domain (APD, purple; including the three-helix bundle, dark purple), the delivery domain (yellow) and the CROP domain (white). b, Cartoon representation of the TcdA1832 structure (coloured according to a), with zinc shown in green. c, The structure in b, rotated 90°, with the GTD shown as a surface view with the UDP–glucose binding site in green. d, The TcdA1832 structure was fit in the 20 Å EM map of TcdA holotoxin (56) using Chimera (217). The InsP6 binding site is shown in green, and positions for the first and last residues visible in the structure are indicated.

The GTD (residues 1–542) is responsible for transferring a glucose from UDP-glucose to the switch I region of Rho-family GTPases and is similar to the structures of the isolated GTDs from TcdA, TcdB and other large glucosylating toxins (Appendix Table 2). In the context of TcdA1832, the GTD is oriented such that the GTPase binding site (proposed based on mutational studies in the TcdB GTD) (129) is occluded by the presence of the APD (Fig. 4-1c). This explains data indicating that glucosyltransferase efficiency is enhanced after the GTD is released by autoprocessing (96). The C-terminus of the GTD emerges in proximity to the APD (residues 543–802), with residues 538–557 forming an extended loop that spans the APD active site (Fig. 4-2a).
Figure 4-2. Zinc is required for autoprocessing activity.

a, The APD, along with a small portion of the GTD and the three-helix bundle from the TcdA_{1832} structure (oriented as in Fig. 4-1a), is depicted with residue 542 in red, residues 543–745 in white, the 746–765 β-flap in light blue, and some of the three-helix bundle (766–801) in dark blue. Zinc (green) is bound in the APD active site by H655, C700 and H759. Four lysines form the initial binding site for InsP6: K602, K649, K754 and K777. b, On comparison with a, the InsP6-bound structure of the TcdA APD (Protein Data Bank: 3HO6) (57) suggests significant structural changes occur with InsP6 binding: the accumulation of eight lysines and one arginine in the InsP6-binding site, a rearrangement of the β-flap and elements of the three-helix bundle, and displacement of H759 from the active site. c, Mutation of TcdA His759 or TcdB His757 leads to proteins that undergo autoprocessing at lower concentrations of InsP6. Cleaved GTD was quantified relative to the holotoxin in the 0 mM InsP6 control, and means ± s.d. (n = 3) are shown. d, Chelation of zinc through treatment with 10 mM TPEN renders TcdA and TcdB incapable of InsP6-induced autoprocessing. e, Chelation of zinc in mutants that show enhanced autoprocessing renders TcdA and TcdB incapable of InsP6-induced autoprocessing. Experiments in d, e were conducted in the presence of 5% ethanol, a solvent for TPEN. f, Autoprocessing can be restored in TPEN-treated autoprocessing-defective preparations of TcdA and TcdB with the addition of ZnCl₂. Gels in c–f are representative of three independent experiments.
Autoprocessing in TcdA and TcdB has been ascribed to an InsP6-dependent cysteine protease activity that results in cleavage after L542 (L543 for TcdB) and release of the GTD (110,111,114,115). Structures of the isolated TcdA and TcdB APDs have shown that InsP6 binds a positively charged pocket, distal from the active site (54,57,218). Structures in the absence of InsP6 have heretofore been unavailable, but mutational studies have revealed an allosteric switch where InsP6 binding is functionally coupled to the active site through a central ‘β-flap’ structure (Fig. 4-2b) (54,57).

The N-terminal portion (547–741) of the TcdA_{1832} APD (crystallized in the absence of InsP6) aligns to the InsP6-bound APD with an alpha-carbon root-mean squared deviation (r.m.s.d.) of 0.67Å, but the C-terminal portion of the domain is significantly different. The β-flap (residues 746–765) separating the InsP6 binding site and the catalytic dyad (C700 and H655) has rotated ∼90° and the sequence that follows (766–802) is significantly repositioned (Fig. 4-2a,b). One effect of this conformational change is an increase in positively charged residues at the InsP6 binding site. The pocket transitions from four lysine residues in the TcdA_{1832} structure (K602, K649, K754 and K777) to include seven lysines and one arginine in the InsP6 bound structure (Fig. 4-2a,b), and thus provides a mechanistic framework for understanding how an electropositive InsP6 binding site can exist in the absence of InsP6. The largest change is evident in Lys766, as its NZ atom moves 21Å as a result of rearrangements in the β-flap. The change also results in a 19Å movement of H759 out of the active site (comparison of Cβ atoms). Mutation of H759 (or H757 from TcdB) results in a protein whose autoprocessing profile no longer varies with InsP6 concentration (Fig. 4-2c), suggesting that this residue is a key regulator of InsP6-induced allostery in TcdA and TcdB.

Analysis of anomalous signals in our diffraction data revealed a zinc atom, bound at H759 and the catalytic dyad of the APD (Fig. 4-2a and Appendix Fig. 6a). TcdA binds zinc in solution, as indicated by both X-ray absorption spectroscopy (XAS) and inductively coupled plasma-mass spectrometry (ICP-MS) experiments (Appendix Fig. 4b and Appendix Table 3). A zinc atom is present at this site in both TcdA and TcdB, as indicated by ICP-MS analysis of TcdA, TcdB and catalytic dyad mutants (C700A and H655A or C698A and H653A in TcdA and TcdB, respectively) (Appendix Table 3). One interpretation for the observation of zinc bound at this site is that zinc acts as an inhibitor of autoprocessing. InsP6 binding could displace His759 (or TcdB His757), causing the release of zinc and the availability of cysteine as a nucleophile.
However, the addition of InsP6 to TcdA or TcdB did not displace zinc from the active site (Appendix Table 3). Furthermore, the removal of the zinc through chelation with \( N,N,N',N' \)-tetrakis (2-pyridylmethyl) ethylenediamine (TPEN) resulted in a loss of autoprocessing activity that could then be restored through the addition of zinc (Fig. 4-2d,f ). Even the TcdA H759A and TcdB H757A mutants that retain zinc binding (Appendix Table 3) were rendered inactive for autoprocessing through treatment with TPEN (Fig. 4-2e). These experiments indicate that zinc is required for the autoprocessing activity of TcdA and TcdB.

A three-helix bundle (767–841) is located at the GTD-APD interface and serves as a transition into the delivery domain (Fig. 4-3a). The three-helix bundle is followed by a small globular subdomain (850–1025) and then an elongated ‘hydrophobic helical stretch’ containing four \( \alpha \)-helices (1026–1135) that extend to the other end of the molecule. The delivery domain then adopts a series of \( \beta \)-sheet structures as it returns to the base of the APD (Appendix Fig. 7). A search for structural homologues using DALI indicates that the delivery domain structure is unique.

Figure 4-3. The delivery domain provides an extended scaffold for an \( \alpha \)-helical hydrophobic stretch involved in pore formation.
Most of the TcdA<sub>1832</sub> crystal structure (residues 1–1025 and 1136–1802) is depicted as a transparent surface with the GTD in white and the APD in blue. The delivery domain is visible as a cartoon to highlight the three-helix bundle (blue), the globular sub-domain (green), the α-helical hydrophobic stretch (residues 1026–1135, pink) and the β-scaffold (yellow). Residues implicated in TcdB pore formation are shown as orange (108) or red (107) sticks. Representative sequences from the six large clostridial glycosylating toxins were aligned and scored with a Risler matrix according to the extent of sequence variation. Scores are displayed on the TcdA<sub>1832</sub> structure surface with a colour ramp (red, orange, yellow, green, light blue, dark blue) in which strictly conserved residues are coloured red and the most variable residues are coloured dark blue. The most conserved surface region (boxed) is at the end of the α-helical hydrophobic stretch: the 1098–1118 loop and β-hairpin. Within this region, the V1109, N1110 and N1111 residues are notable in their accessibility to solvent. The TcdA<sub>SAS</sub> protein binds cells at levels equivalent to wild type but is impaired in its capacity to glucosylate Rac1. Toxins (10 nM) were applied to HeLa cells and incubated for 3 h at 37 °C. Proteins were separated by SDS-PAGE and probed with antibodies that recognize TcdA CROP, non-glucosylated Rac1 or total Rac1. Quantitation of four gels indicates that while 100% of the detectable Rac1 was glucosylated by TcdA, only 23.4 ± 10.8% was glucosylated by TcdA<sub>SAS</sub> (relative to mock treated). The TcdA<sub>SAS</sub> protein is not impaired in its capacity to glucosylate Rac1 in vitro using purified proteins. Toxins (100 nM) were incubated with purified GST-Rac1 for 3 h at 37 °C and analysed as in e. The gel is representative of three independent experiments. The TcdA<sub>SAS</sub> mutant is defective in its cellular toxicity. Toxins (10 fM–20 nM) were incubated with CHO cells for 48 h at 37 °C and viability was normalized to untreated cells. A representative dose–response curve is shown and values of the effective concentration conferring half maximal protection (EC<sub>50</sub>) ± s.d. were calculated from two biological replicates using Prism: TcdA (blue circles, EC<sub>50</sub> = 0.11 ± 0.01 nM); TcdA<sub>DXD</sub> (a glucosyltransferase-defective mutant (219); orange squares, EC<sub>50</sub> < 20 nM); TcdA<sub>SAS</sub>, (purple triangles, EC<sub>50</sub> = 1.74 ± 0.39 nM). Pore formation on biological membranes. TcdA, TcdA<sub>DXD</sub> and TcdA<sub>SAS</sub> were applied to Vero cells preloaded with <sup>86</sup>Rb<sup>+</sup> and then subjected to external medium at pH 4.8. Data represent the means and s.d. associated with four experiments. Colors are as in e.

The hydrophobic sequences in TcdA (958–1130) and TcdB (956–1128) have been predicted to insert into the endosomal membrane with acidic pH to facilitate the translocation of the GTD into the cytosol (73). Residues within this stretch have been shown experimentally to be important for TcdB pore formation, and the corresponding residues in TcdA are highlighted in Fig. 4-3a (107,108). Recognizing that hydrophobic helical elements resemble motifs present in the pore-forming domain of diphtheria toxin (DT) (181,193), we proposed a ‘double-dagger’ model where TcdB inserts two pairs of helical hairpins into the membrane to form a pore in Chapter 2 (107). The pore-forming domain of DT is a globular 10-helix bundle, with the most hydrophobic sequences shielded within the core of the soluble toxin structure, while the helical hydrophobic sequences of TcdA are stretched across the surface of an elongated scaffold of β-sheets. We
propose that this large delivery domain scaffold provides an alternative structural solution to maintaining hydrophobic segments that are destined for the membrane in a soluble, but readily accessible conformation.

In addition to the homology with TcdB, TcdA shares homology with large glucosylating toxins from *C. sordellii* (TcsH and TcsL), *C. novyi* (Tcnα) and *C. perfringens* (TpeL). Sequences from these six large clostridial toxins (LCTs) were aligned and the sequence conservation was mapped onto the TcdA1832 structure (Fig. 4-3b). The largest area of strict conservation that mapped to the surface of the structure was located in a portion of the ‘hydrophobic helical stretch’, a 1096–1115 loop that includes the L1108–N1111 β-hairpin. We mutated the β-hairpin turn from VNN to SAS (a conservative change that maintains two small polar residues). TcdA_{SAS} showed no defect in its cell surface binding (Fig. 4-3c and Appendix Fig. 8) or *in vitro* glucosyltransfer activity (Fig. 4-3d and Fig. 10) but was impaired in its capacity to kill cells (Fig. 4-3e). TcdA_{SAS} is impaired in a cell-surface Rb^{86+} release assay (Fig. 4-3f) and in its capacity to glucosylate Rac1 in a cell-based intoxication assay (Fig. 4-3c), suggesting a defect in endosomal membrane insertion. The identification of a conserved surface turn with essential function in toxicity suggests that antibodies specific for this conserved region could provide protection against multiple toxin-mediated clostridial infections and points to a generalizable strategy for generating safe vaccine antigens for this class of toxins.
Chapter 5

**Functional Defects in Clostridium difficile TcdB Toxin Uptake Identify CSPG4 Receptor Binding Determinants**

The work presented in this chapter was published in the *Journal of Biological Chemistry*.


**Contributions:**

My contributions to this work included designing and analyzing most of the experiments, protein constructs generation, protein expression purification and designing and performing cell biology and biochemical studies for characterization (in collaboration with P. Gupta, Merck & Co., Inc.) P. Gupta performed most of the cell binding and pull-down assays. S. Raman contributed to the gel filtration assays and purification of FZD7 (Julian lab). S. N. Sugiman-Marangos created the illustration of homology model. R. Melnyk and L. D. Hernandez supervised the project.
5 Functional Defects in Clostridium difficile TcdB Toxin Uptake Identify CSPG4 Receptor Binding Determinants

5.1 Overview

Clostridium difficile is a major nosocomial pathogen that produces two exotoxins, TcdA and TcdB, with TcdB thought to be the primary determinant in human disease. TcdA and TcdB are large, multi-domain toxins, each harboring a cytotoxic glucosyltransferase domain that is delivered into the cytosol from endosomes via a translocation domain following receptor-mediated endocytosis of toxins from the cell surface. While there are currently no known receptors for TcdA, three cell-surface receptors for TcdB have been identified: CSPG4, NECTIN3 (PVRL3) and FZD1/2/7. The sites on TcdB that mediate binding to each receptor are not defined. Furthermore, it is not known whether the combined repetitive oligopeptide (CROP) domain, is involved in, or required for receptor binding. Here, in a screen designed to identify sites in TcdB that are essential for target cell intoxication, we identified a region at the junction of the translocation domain and the CROP domain that is implicated in CSPG4 binding. Using a series of C-terminal truncations, we show that the CSPG4 binding site on TcdB extends into the CROP domain, requiring three short repeats for binding and for full toxicity on CSPG4-expressing cells. Consistent with the location of the CSPG4-binding site on TcdB, we show that an anti-TcdB antibody, which binds partially within the first three short repeats, prevents CSPG4 binding to TcdB. In addition to establishing the binding region of CSPG4, this work ascribes, for the first time, a role the TcdB CROPs in receptor binding, and further clarifies the relative roles of receptors in TcdB pathogenesis.

5.2 Introduction

Research over the past decade has provided great insight into the structure and function of the C. difficile toxins, in particular for the individual toxin domains and the key processes that they carry out once inside host cells. Our understanding of how each toxin recognizes and binds target cells, however, is incomplete. Historically, the CROP domain was assumed to be the sole receptor binding domain in the both TcdA and TcdB (71,74), although, the discovery of TpeL from C. perfringens, a homologue of TcdA/TcdB that naturally lacks the CROP domain (220), and the observation that TcdA/TcdB truncations with CROP domains deleted are capable of
intoxicating cells (67,89), have called the role of the CROPs into question. Recent efforts have begun to focus outside the CROP domain to find receptor binding determinants (66,67,108). From these studies, a multiple receptor model for host cell entry has been proposed (67,221). According to this model, it is suggested that toxin docks onto the cell surface by binding to a low affinity receptor/oligosaccharide via its CROP domain, followed by binding to high affinity CROP independent receptor(s), a model suggested for both TcdA and TcdB (13,14).

Recently, three distinct cell-surface receptors for TcdB were identified: the poliovirus receptor like 3 (PVRL3, or NECTIN3), chondroitin sulfate proteoglycan 4 (CSPG4), and members of Frizzled protein family (FZD1, FZD2, and FZD7) (77-79). Remarkably, none of these receptors appear to bind TcdA, despite the substantial sequence identity shared by the toxins. While NECTIN3 has been shown to be important for the necrosis phenotype induced by higher concentrations of TcdB, CSPG4 and FZD have been shown to be important for the cytopathic effects of the toxin that are induced at lower doses of TcdB. That NECTIN3 and FZD proteins were shown to directly interact with TcdB1-1830 indicates that toxin entry via these receptors does not require the presence of the CROP domain (78,79). For CSPG4, the binding determinants on TcdB are not as clear. Based on the ability of CSPG4 to bind TcdB1500-2366 but not TcdB1852-2366, Yuan et al. proposed that CSPG4 is a CROP-independent receptor that binds in a region spanning amino acid residues 1500-1852 (77). Tao et al., on the other hand, proposed that CSPG4 is a CROP-dependent receptor due to lack of binding of CSPG4 to TcdB1-1830 (79). Direct binding of CSPG4 to TcdB1830-2366, however, was not tested.

In this study, we set out initially with the goal of identifying regions in the TcdB delivery domain (amino acids 800-1850)—outside the previously-characterized hydrophobic region (amino acids 956-1128)—that were required for pore formation/translocation. Through this analysis, we identified two residues, Y1824 and N1839, at the junction of the C-terminal end of the translocation domain, and the CROP region, that were essential for functional intoxication by TcdB. Rather than being involved in pore-formation or translocation, however, we discovered that residues in this region were implicated in binding to the TcdB receptor CSPG4. Unexpectedly, CSPG4-binding defective mutants, though still able to bind NECTIN3 and FZD7, showed reduced binding to the surface of cells expressing all three receptors, suggesting that the other TcdB receptors are unable to fully compensate for reductions in CSPG4 binding. Using C-terminal truncations of TcdB and binding of a CROP-targeted antibody, we established that
CSPG4 binding (and full cellular toxicity by TcdB) requires at most the first three short oligopeptide repeats from the CROPs. In addition to identifying the CSPG4 receptor binding determinants, these findings help reconcile previous seemingly contradictory findings about the CROP-dependency for CSPG4 binding. Further, this work provides evidence that the majority of the CROP region, beyond residue 1900, is dispensable for full cellular intoxication by TcdB.

5.3 Materials and Methods

5.3.1 Production of TcdB proteins

5.3.1.1 Cloning and mutagenesis

Plasmids for recombinant expression of TcdB, TcdB \textsuperscript{1-2034}, and TcdB \textsuperscript{1-1900} were generated by Gibson cloning methodology. DNA fragments encoding full length TcdB and the C-terminal truncations were amplified by PCR using genomic DNA from \textit{C. difficile} 10463 strain as the template. The amplified fragments were cloned in frame with the 6X His-tag in the plasmid pHis1522 (MoBiTec, Cat# BMEG10) using the Gibson cloning kit (New England Biolabs, Cat# E2611). Cloning of plasmids expressing TcdB\textsubscript{1-1834} (78) TcdB\textsubscript{1834-2366} (84) and TcdA\textsubscript{1832-2710} (85) was previously described. Single point mutations were made in the TcdB codon-optimized sequence using QuikChange lightning multimutagenesis kit (Agilent Technologies). Plasmids with correct mutations were transformed and expressed using the same conditions as wild type.

5.3.2 Cell Culture and generation of cell lines

5.3.2.1 Cell lines and reagents

Vero, Caco-2, HT29, 18Co, HEK293 and CHO-K1 cell lines were purchased from the American Type Culture Collection (ATCC). All cell lines were grown at 37°C in 5% CO\textsubscript{2}. Vero and 18Co cells were maintained in Eagle’s minimal essential medium (MEM) supplemented with 10% fetal bovine serum (FBS), 100 U/ml penicillin, and 100 U/ml streptomycin. Caco-2 cells were maintained in MEM supplemented with 10% FBS, 100 U/ml penicillin, 100 U/ml streptomycin, non-essential amino acids and 0.75% sodium bicarbonate. HT29 cells were cultured in McCoy's 5A Modified Medium supplemented with 10% fetal bovine serum, 2 mM glutamine, 0.75% sodium bicarbonate, 100 U/mL penicillin, and 100 U/mL streptomycin. HEK 293 cells were cultured in Dulbecco's minimal essential medium (DMEM) (ATCC) containing 10% FBS, with penicillin (100 units/ml), streptomycin (100 units/ml). Chinese hamster ovary cells CHO-K1
cells were cultured in Ham’s F-12 medium (Wisent) with 10% FBS and 100 U/ml penicillin, and 100 U/ml streptomycin. The *C. difficile* VPI 10463 strain (ribotype 087) was purchased from the ATCC. Toxin fragments were expressed in *Bacillus megaterium*. All other chemicals and reagents were purchased from ThermoFisher Scientific unless otherwise stated.

### 5.3.2.2 Generation of CSPG4 expressing HEK 293 Cell Line

pCMV6-CSPG4 vector expressing full length Myc-DDK-tagged CSPG4 was purchased from OriGene Technologies (Cat# RC218462). HEK-293 cells were transfected with pCMV6-CSPG4 plasmid using Lipofectamine-2000 according to manufacturer's protocol (ThermoFisher Scientific). Two days after transfection, the medium was replaced with DMEM supplemented with 10% FBS, penicillin (100 units/ml), streptomycin (100 units/ml), and G418 (1µg/ml). G418-resistant cells were pooled, expanded, and then analyzed for expression of CSPG4 expression by Western blotting with anti-DDK antibody (OriGene Technologies, Cat# TA5011-100).

### 5.3.2.3 CRISPR mutagenesis

Vero CSPG4 knock-out cells were generated via CRISPR-Cas9 technology using Geneart CRISPR Nuclease (CD4 Enrichment) Vector Kit from ThermoFisher Scientific (Cat# A21175). Briefly, two complementary oligonucleotides (5’AACGCCTCCTCGCAGTCCC GGTTTGTTTTT-3’, 5’GGGACTGCAGAGGAGGCGTTCGGTG-3’) encoding a guide RNA fragment were used (Sigma). The protospacer sequence (5’-AACGCCTCCTCGCAGTCCC-3’) was complementary to a sequence in exon 3 of the CSPG4 gene. The two oligonucleotides were annealed and cloned into CRISPR nuclease CD4 vector supplied with the kit according to manufacturer’s instructions and the resulting plasmid was transfected into Vero cells using Lipofectamine 2000. Transfected cells were first enriched using Dynabeads CD4 magnetic beads (ThermoFisher Scientific, Cat# 11331D) and then treated with TcdB for the selection of toxin resistant Vero CSPG4 knock-out cells. A single CSPG4 knock-out clone was isolated by limited dilution. Loss of CSPG4 expression was confirmed by Western blotting and a base pair insertion in the exon 3 of the CSPG4 gene was confirmed by Sanger sequencing of the region targeted by the guide RNA.
5.3.3 Characterization of protein constructs

5.3.3.1 Toxin cell death assay

Effect of toxin constructs on cell survival was assessed as previously described using the sulforhodamine B (SRB) assay (84,222). Briefly, serial dilutions of toxins in cell culture medium were added to cells previously grown overnight in 96-well plates. Following a 24 h incubation at 37°C, cells were washed with phosphate buffered saline and allowed to grow for further 48 h. Cells were fixed by adding 10% cold TCA, followed by incubation for 60 min at 4°C. TCA was removed, wells were washed four times with distilled water, and 100 µl/well of 2 mg/ml sulforhodamine B (Sigma) in 1% acetic acid was added. Plates were incubated for 20 min at room temperature and then washed four times with 1% acetic acid and air-dried. 150 µl/well of 10 mM Tris was added, and plates were incubated with shaking at room temperature for an additional 10 min. Plates were read in a SpectraMax plate reader (Molecular Biosystems) at an absorbance wavelength of 570 nm.

To test for competition between the CROP domain and the toxin, purified toxin was diluted in the culture medium to a final concentration that results in ~90% decrease in cell viability, in the absence and presence of various concentrations of the CROP domain, and then added to cells. After 2 h of incubation at 37°C, cells were washed with PBS, allowed to grow for further 48 h and processed as described above.

5.3.4 Binding Assays

5.3.4.1 TcdB cell surface binding assay

Binding of TcdB to the Vero cell surface was assessed as described previously (84). Briefly, 10-cm dishes of confluent Vero cells were pre-chilled on ice. 200 ng/ml TcdB or the mutants in Vero cell culture medium were added to the cells. Plates were incubated on ice to allow toxin binding. After 30 min, plates were washed three times with cold PBS, and cells were harvested by scraping. Membrane proteins were isolated in the cold using the Mem-PER Plus membrane protein extraction kit (ThermoFisher Scientific, Cat# 89842) according to the manufacturer’s instructions. In the final step, membrane proteins were solubilized in a total volume of 100 µl of solubilization buffer and analyzed by immunoblotting using bezlotoxumab to detect TcdB and
rabbit anti-cadherin polyclonal antibody (Cell Signaling, Cat# 4068) to ensure equal amounts of protein were loaded in each lane.

5.3.4.2 Co-immunoprecipitation assays

Cell lysates for immunoblotting and co-immunoprecipitation studies were prepared by lysing five million cells in 300 µl of NP-40 lysis buffer (ThermoFisher Scientific, Cat# FNN0021) by incubating on ice for 30 min. For studying interaction of TcdB with CSPG4, 3 µg of TcdB and its fragments were incubated with 150 µg of total cells lysate from HEK 293 cells overexpressing myc-DDK-tagged CSPG4 or containing empty vector in 20 mM Tris (pH 7.4), 150 mM NaCl at 4°C for 2 h. TcdB: CSPG4 complex was immunoprecipitated using anti-FLAG M2 magnetic beads using the manufacturer’s protocol (Sigma, Cat# M8823). Briefly, proteins mixtures were incubated with 50 µl of beads at 4°C for 1 h. Beads were washed 3 times with 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.15% Triton-X100. Flag-tagged proteins were eluted by incubating the beads with 100 µl of 150 ng/µl 3X FLAG peptide (Sigma, Cat# F4799) at 15°C for 30 min. For studying TcdB: NECTIN3 interaction, 3 µg of his-tagged TcdB (WT and the mutants) was incubated with 5 ug of His-tagged NECTIN3 ectodomain, a.a. 1-400 (ThermoFisher Scientific, Cat# 10852-H08H-25) in 20 mMTris (pH 7.4), 150 mM NaCl at 4°C for 16 h. TcdB: NECTIN3 complex was immunoprecipitated using 1G10 antibody, targeting TcdB CROP domain (Babcock et al.), coupled to protein A magnetic beads (ThermoFisher Scientific, Cat# 10001D). 10 ug of antibody was incubated with 50 µl Protein A beads in 200 µl of PBS with 0.02% Tween-20 for 10 min at room temperature. Beads were washed to remove excess unbound antibody and incubated with TcdB: NECTIN3 protein complex at 4°C for 1 h. Following 3 washes with PBS + 0.02% Tween-20, proteins were eluted by resuspending the beads in 100 µl of Laemllli sample buffer. Eluted proteins were subjected to immunoblotting analysis.

5.3.4.3 Immunoblotting

Proteins were resuspended in Laemmli sample buffer, boiled, and separated by SDS-gel electrophoresis on 4-12% NuPage Tris-Glycine gels. Proteins were transferred to a nitrocellulose membrane using the iBLOT blotting system from ThermoFisher Scientific. The nitrocellulose membrane containing transferred protein was blocked in Odyssey blocking buffer (LI-COR Biosciences, Cat# 927-40000) for 30 min followed by incubation with appropriate primary antibody for 1 hour at room temperature. TcdB and NECTIN3 proteins were detected using anti-
His antibody (Cell Signaling, Cat# 2365), CSPG4 with anti-FLAG (OriGene Technologies, Cat# TA5011-100). After washing, the blot was incubated with a secondary IgG antibody coupled to IRDye 800CW or 680RD (LI-COR Biosciences) for 30 min at room temperature. After additional washing, bands were visualized using the Odyssey Imaging System (LI-COR Biosciences).

5.3.4.4 Fzd7 purification

The gene for the CRD construct of human Fzd7 (residues 42–179) with a C-terminal monoVenus tag followed by a tandem His12x tag was codon-optimized for expression in human cells (Life Technologies) and cloned into the pHLsec vector. Human Fzd7-CRD was expressed in HEK 293F cells and purified using Ni-NTA affinity chromatography. The protein was eluted with an increasing gradient of imidazole with a maximum concentration of 500 mM, followed by gel-filtration chromatography (Superdex 200 Increase, GE Healthcare) in 20 mM Tris pH 8.0 and 150 mM NaCl buffer.

5.3.4.5 Gel filtration assay

Purified hFzd7-CRD mVenus was mixed with wild type and TcdB mutants at 1:5 molar ratios of TcdB: Fzd7 and incubated at room temperature for 30 min. The mixture was run on a Superose 6 Increase 10/300 GL column (GE Healthcare) equilibrated in 20 mM Tris pH 8.0 and 150 mM NaCl. Eluting peaks were monitored by absorbance at 280 nm.

5.4 Results

5.4.1 Mutations at the boundary of the translocation and the CROPs domain affect TcdB function

As part of our ongoing efforts to elucidate the mechanism by which TcdB intoxicates host cells, we recently developed a platform that enables rapid generation and screening of site-specific perturbations in TcdB (107). We used this platform here to interrogate a poorly defined region of TcdB encompassing the junction of the translocation domain and the CROP domain (Fig. 5-1A).
Figure 5-1. Identification of functional defects at the boundary of the CROP and translocation domain.

**a**, Schematic drawing of TcdB organized into four domains: N-terminal glucosyltransferase domain (red), autoprotease domain (blue), translocation/pore formation domain (yellow) and the C-terminal CROPs domain. The box represents the hydrophobic region at the junction under investigation (N1820-F1856). **b**, Effects of Cys and Lys substitutions on cellular toxicity. Viability of each Cys and Lys mutant was tested by exposing titrated mutant soluble lysates onto Vero cells (three fold titration, 3 days) and quantified by measuring PrestoBlue fluorescence assay. Fold shift of mutant to wild type was calculated by dividing the half-maximum effective concentration (EC50) of mutants by the EC50 of wild type. n=3, ***P< 0.001, two-way analysis of variance (ANOVA). **c**, The cell viability titration curve of wild type TcdB and mutants, Y1824K and N1839K. Toxicity of mutant and wild type toxins on Vero cells were quantified by titrating purified proteins onto Vero cells (four fold titration) in 96-well plates and incubated for 24 h at 37°C (n=3). Forty-eight hours later, the cell viability of treated cells was quantitated by sulforhodamine B (SRB staining).
Despite a wealth of structures available for various TcdA and TcdB fragments containing either the translocation domain (223) or the CROP (84,224), none have provided any structural information for residues between 1803-1834. Based on the prevalence of hydrophobic residues in this region, between 1823-1845 [FYINNGFMMMVSLYINDSLYYF], we initially posited that this region might insert into the membrane during the events in the endosome that involve acid-induced unfolding and formation of the translocation pore (107). Cysteine- and lysine-substitutions were introduced at several conserved residues (across clostridia toxins) and screened, initially from E.coli lysates, for their ability to intoxicate Vero cells, which are widely used for assessing TcdB activity due to their high sensitivity to this toxin. Through this analysis, we identified two sets of mutations, Y1824C/K and N1839C/K, which significantly decreased the ability of these toxins to intoxicate CHO cells (Fig. 5-1B). Purified Y1824K and N1839K toxins were 120-fold and 360-fold less toxic than WT TcdB, respectively (Fig. 5-1C). Surprisingly, other intervening residues had no impact on toxin activity, arguing against the notion that this entire region was inserting into the membrane.

5.4.2 Defective mutants show decreased host cell-surface binding

To establish the mechanistic basis for the observed defects in activity for mutants, we employed a series of assays analyzing Rac-1 glucosylation (glucosyltransferase activity), autoproteolytic cleavage, pore formation, pH-dependent unfolding, and cell surface binding activities. Both defective mutants were equally active as WT toxin in Rac-glucosylation and autoprolytically cleavage assays, and each showed a similar pattern to WT TcdB in the pH-dependent unfolding assay (Appendix Fig. 10A-D) demonstrating the toxins were functional and not misfolded. We next measured the pore-formation functionality of the Y1824K and N1839K mutants using the standard rubidium-release assay (Fig. 5-2A) (107).
Figure 5.2. Characterization of defective purified TcdB mutants Y1824K and N1839K.

a, Pore formation of purified mutant toxins was tested on CHO cells preloaded with $^{86}\text{Rb}^+$ . Pore formation was induced by acidification of the external medium (control pH 7.5; black bars, pH 4.5; grey bars) n=4. b, Relative gene expression of TcdB receptors. The relative gene expression of identified TcdB receptors, NECTIN3, CSPG4 and FZDs on different mammalian cell lines was assessed using a $\Delta\Delta C_q$ method determined from qPCR data. (n=2). Vero cells that were used
in the cell viability assay are highlighted in the red box. e, Cell surface binding of defective mutant toxins to target cells. Immunoblot analysis of TcdB wild type and mutants bound to Vero cell surface at 4°C. The cells were exposed to 200 ng/ml of TcdB for 30 min before being lysed for immunoblot analysis. Membrane bound proteins were detected by anti-TcdB antibody and rabbit anti-cadherin antibody as loading control.

Consistent with the atypical pattern of sensitivity to mutants highlighted above, Y1824K and N1839K were minimally defective in pore-formation. By comparison, defective point mutations in the hydrophobic region of the translocation domain that decreased intoxication efficiency by greater than 10-fold, such as L1106K, gave major defects in pore-formation in this assay (i.e., >90%) (107).

Finally, we measured the ability of defective mutant toxins to bind target cells. To examine this, we investigated their binding to Vero cells, which express all three TcdB protein receptors, NECTIN3 (78), CSPG4 (77) and FZD2 (79), though predominantly CSPG4 (Fig. 5-2B). Cells were incubated in the cold in the presence of toxin to allow binding. Following extensive washing to remove unbound toxin, membrane fractions were then isolated and probed for toxin by western blotting. As shown in Fig. 2C, Y1824K and N1839K were impaired in their ability to bind to Vero cells compared to WT TcdB and the pore-formation defective mutant L1106K (107). Rather than being involved in formation of the translocation pore, these findings suggest that Y1824 and N1839 are involved in binding to target cells. We next undertook to decipher which receptor(s) were involved in binding to this region.

5.4.3 Defective TcdB mutants decrease binding to CSPG4, but not NECTIN3 or FZD7

To elucidate the molecular cause of the defective cell surface binding observed for the Y1824K and N1839K mutants, we tested binding of the mutant toxins to the ectodomains of NECTIN3 and FZD7, a member of the FZD family shown to act as a receptor for TcdB (79), as well as full length CSPG4. By co-immunoprecipitation analysis, WT TcdB, Y1824K and N1839K toxins were able to interact equally with the extracellular domain of NECTIN3 (Fig. 5-3A). Similarly, mutant toxins formed stable complexes with FZD7 as measured by gel filtration demonstrating effective binding to the receptor (Fig. 5-3B). These findings are consistent with previous work indicating that NECTIN3 and FZD proteins bind upstream of TcdB residue 1830 (67,79).
Figure 5-3. Characterization of receptor binding of defective mutants.

a. Association of ectodomain of NECTIN3 with TcdB wild type and mutants, assayed by co-immunoprecipitation analysis. His-tagged TcdB (3ug) was incubated with 5ug of His-tagged NECTIN3 ectodomain at 4°C for 16 hrs. TcdB and NECTIN3 interaction was immunoprecipitated and detected using anti-His antibody. 

b. Interaction of Fzd7 with TcdB by gel filtration assay. Purified Fzd7-CRD mVenus was mixed with TcdB at 1:5 molar ratios of TcdB: Fzd7 and incubated at room temperature for 30min and run on a Superose 6 Increase 10/300 GL column. Elution peaks of only TcdB (black) and the complex with both TcdB and Fzd7 (red) were monitored by absorbance at 280nm, and the complex was visualized on SDS-PAGE.

c. Binding of CSPG4 to TcdB wild type and mutants, assayed by co-immunoprecipitation analysis. TcdB (3ug) was mixed incubated with 150 ug of cell lysate from HEK293 cells.
overexpressing Flag-tagged CSPG4 at 4°C for 2 h. The complex was immunoprecipitated using anti-Flag magnetic beads and visualized by anti-Flag (CSPG4) and anti-His (TcdB) antibodies. 

d, Immunoblot analysis of the wild type (CSPG4+/+) and CSPG4 knock-out (CSPG4−/−) Vero cell lysates for expression of CSPG4 and GAPDH (control). e, Sensitivity of wild type (CSPG4+/+) and CSPG4−/− Vero cells towards TcdB wild type and mutants. Indicated Vero cells were exposed to TcdB (wild type, Y1824K and N1839K) and subjected to cell survival using the SRB assay. f, Toxicity of TcdA wild type and mutants (Y1822K and N1837K) on CHO cells, measured by Prestoblue fluorescence assay as described.

Next, we tested whether the defective mutant toxins were able to bind the CSPG4 receptor. Co-IP studies between mutant toxins and CSPG4, showed that Y1824K and N1839K were defective in binding to CSPG4 compared to WT and L1106K mutant toxins (Fig 5-3C). To confirm this finding, we next generated CSPG4-knockout Vero cells (Vero/CSPG4−/−) using CRISPR-Cas9 technology (Fig 5-3D). Consistent with these residues playing a role in binding to CSPG4, WT TcdB was defective in intoxicating Vero/CSPG4−/− relative to Vero/CSPG4+/+, whereas 1824K and 1839K toxins were both equally active on Vero/CSPG4−/− as on Vero/CSPG4+/+ cells (Fig 5-3E). These data indicate that CSPG4 binding, at least in part, requires determinants that are both in the translocation domain and in the CROP domain, and as such helps reconcile previous studies regarding whether CSPG4 was a CROPs-dependent (79) or CROP-independent receptor (77).

Finally, because CSPG4 is a TcdB-specific receptor, we generated the equivalent mutations to Y1824 and N1839 in the homologous toxin TcdA, which was shown previously to not bind CSPG4 (77). TcdA-1822K and TcdA-1837K showed no defects compared to WT TcdA in intoxicating mammalian cells (Fig. 5-3F), further demonstrating that these residues are involved in CSPG4 binding, and not any other more general aspect of intoxication.

5.4.4 TcdB requires three oligopeptide repeats for CSPG4-binding and full cellular activity

With part of the CSPG4-binding determinants coming from the amino-terminal boundary of the CROP domain of TcdB, we set out to determine how much of the CROP was required for binding to CSPG4. To this end, we engineered, expressed and purified a series of C-terminally truncated toxins of increasing length: TcdB1-1834 (B1834), TcdB1-1900 (B1900), TcdB1-2034 (B2034), TcdB1-2366 (WT TcdB) and a CROP only construct: TcdB1834-2366 (CROP) (Fig. 5-4A).
Figure 5-4. Characterization of TcdB C-terminal truncations.

**a**. Schematic drawings of TcdB C-terminal truncated constructs, B2034 (TcdB<sub>1-2034</sub>), B1900 (TcdB<sub>1-1900</sub>), B1834 (TcdB<sub>1-1834</sub>) and CROP domain (TcdB<sub>1834-2366</sub>). **b**. Interaction between truncated TcdB constructs and CSPG4. His-tagged TcdB constructs were incubated with lysate from cells over-expressing Flag-tagged CSPG4 and subjected to immunoprecipitation by anti-
Flag antibody and visualized by immunoblot as described. c, Sensitivity of Vero and Caco-2 cells towards TcdB C-terminal truncations. Vero cells (left) and Caco-2 cells (right) were exposed to TcdB truncated toxins and cell survival was assessed by the SRB assay as previously described.

To test for an interaction between truncated TcdB constructs and CSPG4, His-tagged TcdB, B1834, B1900, B2034 and the TcdB-CROP were incubated with lysate from cells over-expressing Flag-tagged CSPG4. CSPG4 was then immunoprecipitated with an anti-FLAG antibody and co-immunoprecipitation of the toxin constructs were analyzed by western blotting with an anti-His antibody. As expected, neither B1834, nor the CROP-alone construct interacted with CSPG4 (Fig. 5-4B). By contrast, WT TcdB, B2034 and B1900 were all pulled down with CSPG4, indicating that just three short repeats were sufficient for binding to CSPG4 (Fig. 5-4B). Importantly, the observed binding of truncated toxin to CSPG4 directly correlated with the ability of these toxin truncations to intoxicate CSPG4-expressing Vero cells (Fig. 5-4C). Conversely, in Caco-2 cells, which do not express appreciable levels of CSPG4 (Fig. 2B), B1834 was as potent as the longer TcdB constructs, confirming both that CSPG4 was responsible for the observed defects, and that the truncated toxins were otherwise functional (Fig. 5-4C).

5.4.5 The CROP domain from TcdB is not sufficient for cell surface binding

Our binding data and functional studies with these constructs indicate that the region spanning the junction of the TM and CROP domains is important for binding to CSPG4 but not to the other TcdB receptors NECTIN3 and FZD, which bind upstream of the CROPs (78,79). In addition, since B1834 is equally active as WT TcdB on Caco-2 cells, which do not express CSPG4 (Fig. 5-4C), CSPG4 appears to be the only receptor that requires at least part of the CROP domain for binding. Because of this we questioned whether the TcdB CROP domain by itself has any cell-surface binding activity. To probe this, we analyzed cell surface binding activity of the CROP domains from both TcdA and TcdB. While the TcdA-CROP bound to Vero cells at molar concentrations of 5- and 10-fold greater (15-30 nM) than TcdA (3 nM), no cell binding was detected for TcdB-CROP at up to a 100-fold greater concentration (80 nM) than TcdB (0.8 nM) (Fig. 5-5A and B).
Figure 5-5. Cell surface binding activity of the CROP domains from both TcdA and TcdB.

Vero cells were incubated with TcdA (3nM), TcdA CROP (15nM, 30nM) or TcdB (0.8nM) and TcdB CROP (16nM, 80nM) and surface bound proteins were subjected to immunoblot analysis by an anti-TcdB antibody. Inhibition of TcdA and TcdB cytotoxicity at a fixed concentration by the respective CROP domain in molar excess was measured by SRB assay.
To rule out the possibility that lack of detectable binding was not due to low sensitivity of our binding assay, we carried out a competition assay with the toxins and CROP domains. A fixed amount of either TcdA or TcdB was added to Vero cells in the presence of an increasing concentration of the respective CROP domain. Consistent with the cell surface binding data, the TcdA CROP domain inhibited TcdA activity in a dose dependent manner. However, TcdB activity was unaffected by up to a 20,000-fold molar excess of the TcdB CROP domain (Fig 5-5C and D). These data strongly suggest that the TcdB CROP domain by itself cannot bind to cells and that CSPG4 is the only CROP specific receptor for TcdB.

5.4.6 Antibody binding to TcdB CROPs occludes CSPG4 binding

The anti-TcdB antibody bezlotoxumab, which neutralizes TcdB activity both in vitro and in vivo (83,84,225), was shown recently, using hydrogen-deuterium mass spectrometry and x-ray crystallography, to have two distinct binding sites within the CROP domain of TcdB: E1 and E2, spanning residues 1878-1961 and 2018-2093, respectively (84). A homology model of the segment of TcdB for which there is no structural information (i.e., 1800-1833) was built using sequence and structural alignments via Phyre 2.0 (226). Interestingly, the region that directly precedes the CROPs, 1814-1833, and which contains the defective mutant 1824, models as another short oligopeptide repeat (Fig. 5-6A). An alignment of this region and the downstream CROPs supports this model (Fig. 5-7).
Figure 5-6. Binding model of CSPG4 and bezlotoxumab to TcdB.

a. A homology model of the TcdB segment 1800-1833 with downstream CROP domain was built based on sequence and structural alignment via Phyre 2.0 (yellow) and the crystal structure of CROPs (green) with two Fab domains of bezlotoxumab from Orth et al (84). b, Model of
structure of TcdB segment 1800-1833 and CROP domain with bezolotoxumab binding sites, E1 and E2, and identified key residues mapped to the structure. c, Inhibition of CSPG4 binding to TcdB by bezolotoxumab and actoxumab (control) by immunoprecipitation assay as described previously.

Figure 5-7. Alignment of short and long repeats within the CROP domain of TcdB.
Mapping the defective mutants identified in this study onto this model and structure obtained by Orth et al (84) shows that defective mutants cluster to the same face of the CROPs and in close proximity to the E1 binding site of bezlotoxumab. Given that the CSPG4-binding region established here partially overlaps a portion of E1, we reasoned that bezlotoxumab should prevent binding of CSPG4 to TcdB. To examine this and determine whether bezlotoxumab and CSPG4 binding is mutually exclusive or not, we measured complex formation between full-length TcdB and CSPG4 in the absence and presence of bezlotoxumab and actoxumab (an anti-TcdA antibody used here as a control). Indeed, bezlotoxumab, but not actoxumab, prevented CSPG4 from binding to TcdB (Fig. 5-6).

5.5 Discussion

In this study, we identified a region at the boundary of the translocation domain and the CROP domain that was critical for TcdB intoxication of mammalian cells. Two residues, Y1824 and N1839, flanking a stretch of ~20 largely hydrophobic residues, were highly sensitive to mutation, reducing the ability of TcdB to intoxicate Vero and CHO cells by over two orders of magnitude (Fig. 5-1). Contrary to our initial hypothesis that these residues were involved in pore-formation or translocation, we traced the source of the defects on intoxication to an inability of mutant toxins to bind Vero cells (Fig. 5-2). We showed, using either the ectodomain or full length version of the three known TcdB receptors, that mutant toxins were specifically defective in their ability to bind CSPG4 receptor, but not to NECTIN3 or to FZD7 (Fig. 5-3). Consistent with this, we found that cells lacking CSPG4, either naturally or through targeted removal via CRISPR/Cas9-mediated deletion, were less sensitive to both WT TcdB and TcdB mutants to an equal extent. Along with confirming that mutant toxins are folded and otherwise functional, these data help to further establish that Y1824 and N1839 are involved in binding to CSPG4. Another important observation from these studies is that when present on the surface of a given cell, CSPG4 is the primary receptor for TcdB, since neither FZD7, nor NECTIN3 could fully compensate for its loss.

The location of these two critical CSPG4-binding mutations, at the historically-defined boundary between the translocation domain and the CROPs, prompted us to explore whether more of the CROPs were involved in receptor binding. Using a series of C-terminal truncations, we found that keeping just three short repeats (i.e., B1900) was sufficient for TcdB binding to CSPG4 (Fig.
B1900, missing the majority of the CROPs, is as potent as WT TcdB on Vero cells, indicating that this represents a fully functional form of TcdB on cells expressing CSPG4. These findings suggested that the CROPs play a subtle role, if any at all, in the context of cellular intoxication. This was reinforced by our experiments testing the ability of the TcdA and TcdB CROPs to bind to cells (Fig. 5-5). Whereas the TcdA CROPs were able to bind to the surface of cells, and completely inhibit binding and intoxication by TcdA, the TcdB CROPs showed no evidence of binding to cells, and was unable to competitively inhibit TcdB intoxication, even up to a molar excess of 30,000-fold. These data argue against the notion that TcdB utilizes the CROP domain to dock to the host cell surface by interacting with oligosaccharides followed by binding with specific cellular receptors (67,221). Based on these data alone, however, we cannot exclude the possibility that the CROPs in TcdB beyond 1900 (i.e., 1901-2366) play a different role in the context of an in vivo infection. What is clear from our data here and has been reported previously (66), is that TcdA utilizes its CROP domain to interact with cell surface receptors and that this interaction is essential for TcdA cellular intoxication. These findings further highlight the functional differences between TcdA and TcdB.

The data presented here suggest that the majority of CROP domain in TcdB does not play a role in receptor binding, however, it is well established that neutralizing antibodies, which bind the CROP domain, block TcdB activity both in vitro and in vivo (83,84,222). Our finding that CSPG4 binding required three short repeats of the CROPs that overlap the E1 binding site of bezlotoxumab helps reconcile this apparent discrepancy (Fig. 5-6). Consistent with the proposed binding site, in the presence of bezlotoxumab, CSPG4 can no longer bind full length TcdB, suggesting binding of the antibody interferes with the potential binding pocket of CSPG4. It also supports our finding that the CSPG4 binding site is at the junction between residues 1810-1850, where two key residues are located, highlighting the importance of this junction. Unfortunately, there is no structural information on this junction from existing crystal structural data from both TcdA (TcdA 1-1810) and TcdB CROP domain (1834-2366). Future studies should consider keeping this junction intact for binding and structural analysis.

The notion of defining a “receptor binding domain” for TcdB, let alone receptor-binding domains for each of the individual receptors of TcdB has remained elusive for TcdB. This arises, in part, from issues associated with studying individual domains that have been truncated at previously-defined domain boundaries, which presume a lack of functionality for these
boundaries. Indeed, truncating TcdB at either of the historical boundary sites (i.e., 1852, and 1834), yields two toxin fragments, neither of which can bind CSPG4. Our finding that CSPG4 requires determinants from the translocation domain and the CROPs is satisfying as it helps reconcile the somewhat contradictory conclusions regarding CSPG4’s CROPs dependency (77,79). We opted to not explore the N-terminal boundary of CSPG4 binding in detail beyond what had been done previously (77) since we were most interested in describing the functional determinants of binding, which required upstream factors from TcdB. Nevertheless, from previous work showing that TcdB_{1500-2366} co-precipitates CSPG4, we can minimally define the CSPG4 binding region to within these boundaries. A more detailed description of the binding site likely awaits a high-resolution structure of the complex of the CSPG4 with TcdB, or fragments thereof, which this work will help guide.

Lastly, our findings here also highlight the importance of carefully considering cell lines for studying toxin binding and cytotoxicity analysis. Our gene expression analysis shows very different receptor expression profiles for the different mammalian cell lines that are widely used (Fig. 5-2B); for example, HeLa and Vero express high level of CSPG4 and low level of NECTIN3 while Caco-2 highly expresses NECTIN3 but no CSPG4, and HT-29 expresses both similarly. As shown in our cytotoxicity data, cellular expression of specific receptors could have significant impact on cellular sensitivity to the toxins. Previous findings show that both FZDs and NECTIN3 are found predominantly on colonic epithelial cells whereas CSPG4 is expressed mainly in subepithelial myofibroblast cells like Co-18 cells (78,79,227). In combination, they could serve as targets for different stages of toxin entry. The physiological roles and clinical relevance of all three receptors remain to be elucidated.
Chapter 6

Conclusion, Discussion and Future Directions
6 Conclusion, Discussion and Future Directions

Toxin uptake has been a complicated, enigmatic yet intriguing topic among pore-forming toxins. For my thesis, I focused on *C. difficile* toxins, mainly TcdB, aiming to gain better understanding of the process and help resolve some conflicting results in the field regarding cell surface binding and translocation.

6.1 Mechanism of Pore-formation and Translocation

Through the identification of key residues that were essential for pore-formation and translocation in TcdB, combined with careful hydropathy analysis, I proposed a model for the TcdA and TcdB pore (Fig 2-6). Focusing on the conserved hydrophobic region, I identified highly sensitive residues in the hydrophobic region that were significantly defective in pore formation. These residues mapped to 1037-1109 (TcdA) and 1035-1107 (TcdB). The most defective point mutant, L1106K in TcdB, seems to have tunable toxicity; the polarity of the residue at this position was inversely correlated with membrane-insertion, pore-formation and functional intoxication, confirming L1106 is critical in membrane insertion. The hydropathy patterning of the translocation domain for TcdA and TcdB uncovered an unexpected link to the distant diphtheria toxin. A standard hydropathy plot of the translocation domain of DT revealed two continuous regions of hydrophobicity: a smaller one followed by a larger peak, corresponding to the two helical hairpins that insert into target membranes prior to translocation (Fig 6-1a). Strikingly, a hydropathy plot of the hydrophobic region of TcdA and TcdB exhibited the same general pattern of hydrophobicity (Fig. 6-1b).
Figure 6-1. The link between diphtheria, TcdA and TcdB.

Hydropathy analysis of (a) diphtheria toxin (DT), (b) TcdA (orange, dotted line) and TcdB (orange, solid line) using the membrane protein topology prediction method TMHMM v2.0. Peaks on the hydropathy plot that correspond to transmembrane helices (TH) and hydrophobic helices (HH) for DT, TcdA/TcdB, respectively, are labeled. (c) The pre-pore structure of the DT translocation domain (275-378; PDB 1MDT) and (d) the TcdA/TcdB (1018-1118; 1016-1116; PDB 4R04). The double helical models of the pore for (e) DT and (f) TcdA/TcdB.
The helical hairpins of diphtheria have been extensively characterized as pore forming elements (91, 228-230). For the diphtheria toxin pore, it is believed that transmembrane helices form two \(\alpha\)-helical hairpins, each of which resemble “daggers”, thus invoking the “double-dagger” model for membrane insertion for DT (181). Notably, mutation of the charged residues at the tips of the daggers was found to be the most detrimental for function, likely because they have to travel farthest across the bilayer during insertion (195). With TcdA and TcdB, the clusters of identified highly sensitive residues are located and comprised the “tip” of the daggers (Fig 2-4). Taken together the hydropathy patterning, the similarities to DT and the mutagenesis data, a “double-dagger” similar in principle to DT was proposed for TcdA and TcdB (Fig 2-4). The identified pore-forming region leads to our proposed model similar to DT, making advances in our understanding of translocation but also bringing out more questions, which I will discuss in detail.

6.1.1 The link between C. difficile toxins and other pore-forming toxins

Interestingly, for DT toxin, the most hydrophobic segments are hidden in the core of the pre-pore structure and are assumed to exposed in response to acidification for pore-formation and translocation. The striking similarity in hydropathy between pore forming regions of C. difficile toxins and diphtheria toxin is unexpected, and it is unclear what the connections between these non-homologous toxins are. When the first the connection was discovered, due to lack of high-resolution structural information of TcdA or TcdB translocation domain, we assumed that C. difficile toxins might adopt a similar “bundled” configuration as DT (Fig 6-1). However, the high-resolution structure of TcdA 1-1832 revealed a very unique structure for TcdA translocation domain, with no identified structural homologs, which is different from our original hypothesis (Fig 4-3). The translocation domain, composed of mainly \(\beta\) sheets, adopted an elongated overall structure. The hydrophobic pore-forming region, largely composed of flexible loops and \(\alpha\)-helices in this pre-pore structure, extends from one end of the translocation domain to the ether, wrapping around \(\beta\)-sheet structure (Fig 6-2). Towards the end of the helical region, there is a loop and a \(\beta\)-hairpin, where our highly sensitive mutants are clustered (Fig 6-2).
Figure 6-2. Structure of TcdA, with emphasis on the translocation domain.

(a) The TcdA$_{1-1832}$ holotoxin structure (PDB: 4R04), with the GTD (1-543) in red, the APD (544-799) in blue, the translocation domain (800-1833) in yellow and the hydrophobic region (958-1130) in orange. (b) A closer look at the hydrophobic helical region stretched across the translocation domain. (c) The highly conserved loop and \(\beta\)-hairpin (1096-1115), with sensitive residues to pore formation and translocation highlighted.
The toxin’s strategy to stretch its hydrophobic region across and around the entire translocation domain in an extended conformation rather than bury it is unique among toxins. For instance, diphtheria toxin has an “inside-out” membrane protein fold, where the most hydrophobic helical hairpin is sheltered within a bundle of amphipathic α-helices (181); this fold is present in colicin, a membrane-perforating protein (231) and eukaryotic proteins of BAX and BCL-2 that permeabilize mitochondria (232). Botulinum has a long pair of α-helices with triple helix bundles at either end (182), which bears some resemblance to coiled coil viral proteins, such as HIV-1 gp41/nc4 (233) and influenza hemagglutinin (234), which have ability to penetrate membranes at low pH. The difference in the way the two toxins pack the hydrophobic stretch questions on the link between *C. difficile* toxins and diphtheria toxins. It seems that they adopt different solutions to hide hydrophobic pore forming regions within the prepore structure. It is unclear whether it is due to hydrophobicity of the pore forming regions of the toxins or different mechanism of unfolding in response to low pH. Diphtheria toxin is less hydrophobic than TcdA and TcdB; nevertheless, for pore-forming toxins including TcdA and TcdB, they are still only marginally hydrophobic. The stretch configuration may be the lowest energy configuration in the pre-pore structure while the hydrophobic segments might adopt the diphtheria-like bundles configuration during pore formation (Fig. 6-1).

One of the challenges to study the pore-forming bacterial toxins is that they quickly interchange from water-soluble prepore state to membrane inserting hydrophobic membrane protein state in response to pH, which is generally not stable. It is difficult to study the transient transformation from prepore to pore structure. Even for the better-understood diphtheria toxin and botulinum toxins, the only structural information available is for the prepore state. Structural information on the translocating pore is still lacking. Most of the studies towards the model came from indirect biochemical approaches. For translocating toxins with β-barrel pores, including anthrax and other binary proteins from *Clostridium* and *Bacillus*, pore formation and translocation processes are generally well understood. β-barrel toxins need to oligomerize to become a functional pore, and enzymatic domain must unfold to translocate through the pore, which usually has a narrow constriction in the pore lumen (106). In comparison, there seems to be no agreed upon molecular features or determinants for helical translocating toxins. There is no evidence suggesting that helical pores must oligomerize to accommodate a translocating substrate, which probably unfolds. On the contrary, the molecular feature and determinants for helical translocating pore
remain unresolved in the field. Part of the reason is probably due to the difference between the overall structural stability and interactions between \( \alpha \)-helices in a \( \alpha \)-helical pore and \( \beta \)-sheets in a \( \beta \)-barrel within membranes. Whereas \( \beta \)-barrel pores generally exhibit extreme stability (i.e., the anthrax toxin protective antigen pore is stable in SDS detergent and high heat), owing to extensive inter-strand hydrogen bonding, helix-helix interactions in membranes are rarely SDS-resistant (235,236). It seems that \( \alpha \)-helical pores are more flexible, dynamic and variable in architecture and other features. Indeed, whereas anthrax toxin pore cannot accommodate and translocate highly stable mCherry protein, which is resistant to unfolding, diphtheria toxin is reported to be able to translocate substrates varying in size, stability and structure (237,238). There have been questions regarding whether \( \alpha \)-helical pore is indeed a structured functional translocase or the possibility that translocation occurs simply by disruption of the membrane to create a passage by inserting hydrophobic helices into the membrane while the substrate might not fully unfold to translocate. It is a controversial and heated topic in the field. Due to the similarity in the hydropathy pattern, the same question applies to \( C. \) difficile toxins that whether a functional pore is formed and whether substrate unfolding is required for translocation. Interestingly, the unique flickering behavior for \( C. \) difficile toxins in electrophysiology studies might be an indication that the pore is highly unstable and rapidly changing. Despite the noted similarities between TcdA/TcdB and DT, it is likely that there are many variations for the \( \alpha \)-helical toxins.

6.1.2 Requirements for TcdA and TcdB translocation machinery

Our proposed double-dagger model of the TcdB pore posits that four transmembrane helices comprise the translocation pore. The major pore-forming double-dagger helices in our model comprise only approximately 100 residues from amino acid 985 to 1112. This raises several questions regarding the mechanism of translocating GTD and APD domains given the expected small diameter that a four-helix bundle would create relative to even a fully unfolded protein. Given the discrepancy, it is conceivable that additional and yet unidentified regions might also involve in membrane insertion to form a functional pore. Within the ~1000-residue translocation domain, it seems that at least 100-200 residues in the hydrophobic region are essential for translocation. It begs the question as to what the function of the remaining regions of the translocation domain (800–900 residues) is. As alluded to above, a portion of the C-terminus of translocation domain is contributing to receptor binding, but the exact bounds are not clear, with
several independent studies providing contradictory results. However, these data strongly suggest that the translocation domain is, in fact, a translocation and a receptor binding domain, which complicates the mechanism.

So far, there are two regions within the translocation domain that have been implicated in translocation and pore formation: our proposed hydrophobic helices (roughly around 1018-1112 out of the hydrophobic region 958-1128) and an additional region upstream of the hydrophobic stretch (residues 830-990) (107,108). The first region is the proposed helices that Genisyuerek et al described as the minimal pore-forming region for TcdB (108); partially overlapping with helical hydrophobic stretches, it comprises the globular sub-domain formed by α-helices, as shown by the crystal structure of equivalent segment in TcdA (Fig 6-2). Additionally, they identified within this region an amino acid pair glutamate -970 and glutamate-976, when mutated to lysine, exhibited reduced cytotoxicity and ability to form pores. It has been suggested for pore-forming toxins that pH-dependent membrane insertion of toxins may be dependent on key acidic residues, which are located in a loop between two hydrophobic helices and are important for membrane insertion. These negatively charged residues are protonated and neutralized at the low pH of endosomes, which might be a prerequisite for initial membrane insertion. The idea of pH-dependent sensor has been described in other toxins including DT and cytotoxic necrotizing factor from Escherichia coli (195, 239). They proposed that a similar role for the double mutants. Although we showed that this double mutant was only marginally defective in pore formation and toxicity, it is conceivable that this region also contribute to pore-formation at low pH.

The pore-forming residues we identified from mutagenesis studies located were found clustered to the four α-helices, wrapping around the extended β-sheet structure, following the globular sub-domain, supporting our model that these helices are critical for pore-formation. Within the hydrophobic helical stretch, one of the most critical residues, L1106K, was located at a surface loop covering residues L1108-N1111 in TcdA. This region is conserved in LCTs and contained the two critical residues (L1106 and V1107 in TcdB) that are critical for pore formation and cytotoxicity in both TcdA and TcdB, furthermore, in vivo and in vitro. These findings suggest that this conserved surface loop is key in mediating membrane insertion, providing a great target for designing inhibitors that could block pore-formation and toxicity for general LCTs.
For DT, in addition to the double hairpins, the N-terminus of the translocation domain formed by helices is required for efficient delivery (240). A similar role could apply to *C. difficile* toxins that during pore formation that both the stretched hydrophobic regions formed by two helical hairpins and the N-terminal globular domain might collectively self-assemble to form an initial insertion unit that precedes a fully functional pore for translocation. It remains to be studied that how the hydrophobic stretches respond to acidic conditions to dissociate from the beta-sheet scaffold and whether oligomerization occurs to form a functional pore for translocation. It also begs the question of which region acts as the main driver during insertion. Whereas residues we identified completely abolish toxin function by fully blocking pore-formation, the double mutants were only marginally defective, suggesting that the hydrophobic region is likely composed of four helices and the upstream region which may function as a sensor at early stage.

### 6.1.3 Pore formation vs. Translocation

Studying structural features of pores has been challenging for translocating toxins, due to the marginal hydrophobicity of pore forming elements, and multiple insertion states in the membrane. For TcdA and TcdB, the large size, the ability of many fragments to interact with the membrane and role of receptor binding in the same domain further complicate identification of the key determinants for pore formation. Identifying the translocation machinery instead of pore forming elements may be especially useful tool for TcdA and TcdB, since many regions interact with the membrane, but not all of them are sufficient for translocation. The difficulty for studying translocation is that it is nearly impossible to separate pore formation from translocation. Originally, I had hoped to identify translocation defective mutants that are capable of forming pores but not able to translocate proteins, which would be keystones to elucidating the necessary elements that convert an ion-conducting hole into a working translocase. Each functionally defective mutants identified from our screen, however was impaired in pore-formation, or receptor binding. I was unable to identify a single mutation that was solely defective in translocation. One of the reasons could be the methods I used for screening; $^{86}\text{Rb}^+$ release assay and 48-hour cell viability assay might not be sensitive enough to capture translocation residues. It is possible that the transition from pore formation and translocation is rapid and transient; it would be challenging to independently study translocation. A translocation assay, which can measure the rate of translocation, would be helpful to better understand the process.
Very recently, our collaborator Dr. Lacy reported a neutralizing antibody PA41 against TcdB (241). Previously, monoclonal antibodies have been shown to neutralize *C. difficile* toxins by targeting cell surface binding. The mechanism of this antibody PA41 is very unique in that it blocks pore formation or translocation. The antibody does not inhibit cell binding, cell entry, or enzymatic activities of TcdB; however, PA41 significantly reduces both TcdB-dependent rubidium release on biological membranes (Figure 6-3A) and the release of free GTD into cytosol (Figure 6-3B, C).

Figure 6-3. PA41 inhibits pore formation and the subsequent delivery of the GTD into the cytosol.

(A). Pore formation on biological membranes. TcdB holotoxin alone (1nM) or preincubated with PA41 mAb or Fab (10nM) was applied to CHO-K1 cells proloaded with $^{86}$Rb$, and rubidium release was compared at pH 7.5 and pH 4.5. Means ± SD (n=6) are shown. (B) Caco-2 monolayers were intoxicated with 3xFLAG-TcdB (25 nM) preincubated with equimolar (25 nM) amounts of control or PA41 mAb. Toxin cleavage assays were performed as described in Methods. Blots were probed with antibodies against the toxin (anti-FLAG; detects both internalized holotoxin and free GTD), unglucosylated and total Rac1, and GAPDH. (C) The fraction of the internalized toxin that is cleaved and released in each condition was determined by normalizing the free GTD signal to the corresponding internalized holotoxin signal. For each time point, the fractional GTD release measurements were normalized to that of the isotype mAb controls to obtain the relative GTD cleavage. Means ± SD (n=3) are shown.
It is surprising that this antibody, which blocks pore-formation and translocation actually binds outside of the translocation domain, instead interacting directly with GTD. The exact molecular mechanism of PA41 neutralization is not clear; however, as GTD is not required for pore-formation, we can speculate that binding of PA41 to GTD prevents translocation specifically. The blockage of translocation could be due to either binding of PA41 prevent GTD unfolding for translocation or presence of bound PA41 blocks translocation by occlusion. Understanding the mechanism of action of how PA41 inhibits translocation will be helpful in understanding this process and answer questions like whether unfolding of cargo is required for translocation and nature of the translocating pore. It is a first antibody that targets the discrete mechanistic step of translocation, providing a useful tool to study the process separated from pore formation. Furthermore, given the complexity of multiple receptor model, it provides additional and novel strategy to neutralize toxicity by targeting the more generalized translocation and pore formation processes to treat and prevent CDI.

6.1.4 Future directions

Future directions primarily revolve around verifying and refining the double-dagger model of insertion for TcdA and TcdB by continuing to identify determinants of insertion, pore-formation and translocation and structural features that enable these processes. Understanding these processes for TcdA and TcdB would also help to understand the general mechanism for bacterial toxin translocation. Also, it appears for TcdA and TcdB that receptor binding and pore-formation/translocation are the only steps that are absolutely required for cytotoxicity. It has been shown that inactivation of GTD or APD cannot fully block toxin function but pore-formation is essential for both GTD-dependent apoptotic cell death and GTD-independent necrotic cell death. Thus, elucidation these processes, especially pore formation/translocation, would be crucial to understand toxin function and also provide a generalized strategy to target and block C. difficile toxicity or even apply to LCTs.

Towards understanding translocation, it is important to identify additional regions or segments that might also be involved in pore formation. The strategy employed in the initial mutagenesis screen was biased towards absolutely conserved residues that were within the hydrophobic region. It would be useful to carry out an unbiased screen and lower the standard from 100% conserved residues to highly conserved residues to expand our library. As key residues might not
be absolutely conserved among LCTs and could be located in additional regions that are marginally hydrophobic, some critical determinants might be filtered during the first screen. Given the size of the domain, unbiased screening of the entire domain can be challenging and time consuming. To optimize the process, putative membrane-insertion regions can be screened in segments. Insertion into biological membranes could be evaluated using an expression system that permits quantitative assessment of the insertion efficiency of natural and designed segments using the leader peptidase or Lep system. Segments of interest are engineered into the luminal P2 domain of the integral membrane protein leader peptidase and insertion efficiency is assessed by integration into membranes of dog pancreas rough microsomes (RMs) (242,243). Because glycosylation of the engineered Asn-X-Ser glycosylation sites can occur only in the lumen of the RMs, TM insertion could be distinguished from secretion by simple gel assays. In preliminary studies, I found that putative helix 4 and 5 inserted into membranes with 50% and 20% efficiency, respectively. Interestingly, as a hairpin, the insertion efficiency increased to 90%, suggesting a degree of cooperatively. Using this method, one could systematically evaluate the entire translocation domain by designing 20-residue segments with overlapping segments of 10 residues to identify additional membrane insertion segments.

Following identification of additional key determinants, a library of cysteine mutants could be generated to characterize the structure of the pore domain and of the conformational changes that take place during pore-formation and translocation using site-specific labeling approaches. Based on the high-resolution structure of TcdA translocation domain and identified key residues, distance measurements between specific sites labeled with paramagnetic probes using electron paramagnetic resonance (EPR) spectroscopy could be used to probe the local environment during various states. One could utilize EPR to map conformational changes during translocation by measuring spin-label mobility, and spin-spin interactions between spin-labels placed at two positions within the toxin. This could help to answer questions regarding the determinants for helical segments to unfold from the scaffold and whether oligomerization occurs for C. difficile toxins. Using a combination of biophysics and cell biology, this could help to uncover the mechanism by which a translocation domain of a prototypic bacterial toxin inserts into membranes, assembles into pores and transports the enzymatic domain into cells and help define key principles that underlie the general processes of membrane insertion, protein unfolding and protein translocation for a typical α-helical translocase.
6.2 Receptor binding

6.2.1 Role of CROP domain in TcdA and TcdB

It is clear that the domain historically identified as the receptor binding domain, the CROP, does not bind all, or perhaps any, of the receptors identified to date, and thus, does not constitute the entire receptor binding domain. The TcdB receptors FZD and PVRL3 bind to the C-terminus of the translocation domain, while CSPG4 binds to the junction between the translocation domain and the CROP. We have shown that only a small segment from the CROP is required for full toxin function in vivo, while the remaining region of CROP is dispensable. If only a small portion of the CROP is involved in receptor binding of TcdB, what is the function of the remaining repeats of the CROP? It has been proposed that the CROP domain of TcdA and TcdB might mediate non-specific binding to the cell surface via a low-affinity receptor or via carbohydrates, which could then facilitate binding of CROP independent receptors (67,221).

Importantly, no carbohydrates have been identified that bind to TcdB with high affinity, and evidence of carbohydrates mediating receptor binding for TcdB has not been shown. In our study, we showed that CROP from TcdB did not bind to cell surface and was unable to compete with full-length toxin, suggesting that CROP domain alone is sufficient for binding. For TcdA, the CROP domain indeed contributes to cell surface binding, but it is unknown if the CROP domain from TcdA interacts with specific receptors other than non-specifically binding to carbohydrates. Although there has not been work that directly reveals the function of the CROPs, they do seem important, as targeting both the TcdA and TcdB CROPs with neutralizing antibodies greatly attenuates cytotoxicity (84).

While our a priori assumption is that TcdA and TcdB CROPs have a similar role in pathogenesis, it may be worth considering an alternate hypothesis. Namely, do TcdA and TcdB CROPs have different roles in pathogenesis? This hypothesis is not unwarranted, as the CROP differ in size (TcdA CROP is ~900 residues, TcdB CROP is 600 residues) and overall structure (58,59,63,244). Several lines of evidence suggest that the TcdA CROP may play a role in overall protein folding, and even in interacting with an N-terminal region of the toxin to prevent premature autoprocessing (117,118,164). Binding to the central and C-terminal regions of the TcdB CROP by the neutralizing antibody acetoxumab significantly alters the conformation of the whole toxin and causes aggregation, further substantiating the importance of TcdA CROP in stabilizing the
toxin (85). Overall, while the CROP domain is important but not necessary for toxin function, they may play different roles in TcdA and TcdB. We cannot rule out the possibility that the CROP domains play additional roles in pathogenesis, which needs to be investigated by *in vivo* studies.

6.2.2 What is the physiological relevance of the multiple receptor model?

Identification of at least three receptors for TcdB may explain the past difficulty in identifying inhibitors to block TcdB entry into cells. TcdB may utilize multiple receptors with different binding site to broaden the selection of mammalian cells it can target. Both PVRL3 and FZDs are highly expressed on surface epithelium of the human colon, while CSPG4 is predominantly expressed in the multi-nucleated intestinal sub-epithelial myofibroblasts (ISEMFs) (227). The expression of PVRL3 and FZDs in colonic epithelium suggests that PVRL3 and FZDs might be the first receptors TcdB encounters when released into the lumen of colon. PVRL3 specifically mediates cytotoxic effects of necrosis, which are independent of glucosylation but related to release of reactive oxygen species (164,194). On the contrary, FZDs are responsible for toxin-induced cytopathic effects and cytotoxic cell death by apoptosis under low concentration of TcdB in the picomolar range, similar to CSPG4 (245). Both receptor mediated effects might be important in the disease paradigm, with cytopathic effects first leading to inflammation and cell junction disruption, and necrotic effects later leading to damage of colonic tissue in severe cases of CDI. In addition to its role as receptor for TcdB, frizzled proteins are receptors for Wnt signaling pathway, which is essential in maintaining colonic stem cells. Healthy colonic stem cells constantly supply new colonic epithelial cells, which is key in colonic epithelial cell renewal and repair (246). TcdB competes with Wnt for binding to FZDs, and subsequently, inhibits the Wnt signaling pathway. This suggests that colonic stem cells are a potential target in *C. difficile* pathogenesis (245). Even though CSPG4 is not expressed in colonic epithelial cells, expression of CSPG4 on ISEMFs might serve as an important target to cause further tissue damage after PVRL3 and FZDs mediated disruption of colonic epithelium by exposing sub-epithelial myofibroblast cells. The physiological role and clinical relevance of all three receptors in disease pathogenesis requires further investigation.
6.2.3 Future directions: elucidating TcdB receptor binding sites

Our work identified a clear functional C-terminal boundary for TcdB that is sufficient for toxin function. Only three short repeats from CROPs are essential for CSPG4. However, we were unable to identify the N-terminal bounds for TcdB due to the limitation of the functional assay. Our identification helped further refine the binding region for three receptors; FZD and PVRL3 both bind upstream of residue 1830 while CSPG4 requires the junction between translocation domain and CROPs to bind. The multiple receptor model for TcdB highlights the complexity to study *C. difficile* toxin binding and translocation. It is not clear what the relative role of each receptor and their interaction with each other is. It seems that CSPG4 and FZD do not compete with each other for binding (79), suggesting they have independent non-overlapping binding sites, which is supported by the evidence that one is partially CROP-dependent while the other is a CROP-independent receptor. It remains unclear that whether PVRL3 and FZD compete for binding as they all bind upstream of CROP domain and are commonly found in colonic epithelial cells. It is essential to further define the specific binding region for all three receptors for developing specific inhibitors blocking cell surface binding. It would also help to delineate and redefine the true translocation domain, as it seems like a great portion of the domain is, in fact, contributing to receptor binding, which complicates the studies for translocation and pore-formation. A future direction for the field must be to further refine and elucidate binding site for the three TcdB receptors. Aside from obtaining structures of toxin-receptor complexes, the most straightforward method is through protein truncations, which I have made several for identifying C-terminal boundary for TcdB binding. To simplify the complexity of multiple receptor model working together, by generating individual receptors knockout mammalian cell lines by using CRISPR/Cas9 method, the receptor binding and the impact on receptor mediated cell response can be individually studied.

The challenge I encountered with generating truncations for TcdA and TcdB toxins was that protein truncations and deletions are susceptible to misfolding, which is difficult to evaluate as N-terminal truncations lack functional enzymatic domain. Given the structural of translocation domain from TcdA, majority of the domain is formed by β-sheets and act as scaffold stabilizing the hydrophobic region; it is reasonable to expect making truncations into the scaffold could significantly impact the folding and overall stability of the protein. One needs to take precautions when generating truncations within this domain. Structural studies will be
particularly helpful to identify the exact binding site for TcdB receptors. We are working with a collaborator to see whether we can identify the binding site by crystallization of TcdB C-terminal fragments in complex with TcdB receptor proteins.

6.2.4 Identifying Receptor binding for TcdA

With multiple receptors identified for TcdB, the pathophysiologically relevant receptors of TcdA are still unknown. Although several studies showed that CROP from TcdA indeed contributes to cell surface binding, the multiple receptor model may also apply for TcdA. Truncated TcdA lacking part of the CROP domain can still bind, enter and intoxicate cells, although with reduced efficiency (66). Similarly, it appears that TcdA CROP contributes to but might not be that critical for host cell binding. However, the fact that CROP from TcdA is able to bind cell surface alone and compete with full length toxin for binding, indicates that CROP domain from TcdA, indeed, plays a much more dominant role in cell binding compared to TcdB. Interestingly, studies reported that TcdA comprising residue 1361-1874 is capable of binding (247); furthermore, full length TcdA was shown to have no impact on TcdA C-terminal fragment 1-1874 binding to host cell in competition assay (66). These findings imply that full length TcdA with intact CROP does not bind to the same receptor as the truncated TcdA lacking part of the CROP domain. There might be alternative receptors for TcdA that are shielded by the presence of full CROP domain. Because CROP from TcdA is shown to repress APD activity due to inter-domain interactions and also stabilize the overall toxin (94), it is conceivable that the interaction with alternative receptor is likewise blocked until conformational change exposes the additional binding sites. It may also explain why despite their high sequence conservation and structural similarity, specific receptor remains unidentified for TcdA; likewise, none of the TcdB receptors applies to TcdA. TcdA CROP domain adopts an extended S-shaped structure, which is considerably longer than the horseshoe-shaped structure for the shorter TcdB CROP domain. To identify whether there is a shielding effect from TcdA extended CROP, it is reasonable to generate both full length and TcdA truncations with partial CROP domain (1-1875, 1-1832) to carry out gene knock out screening to identify receptors for TcdA. To test whether TcdA requires the complete CROP domain for complete function, one could make corresponding C-terminal truncations in TcdA and test the impact on binding and toxin function. These experiments can provide insights into potential receptors for TcdA inside or outside of TcdA CROP domain and additional binding sites for TcdA.
6.3 The BioID method to elucidate binding and translocation

All of the above results highlight the complexity of toxin B binding and entry into host cells, suggesting that there could be multiple receptors that interact with multiple binding sites inside and outside the CROP domain. To further discover potential receptors and their binding sites for TcdA, we plan to apply the Bio-ID method for elucidating TcdB and receptor interactions. The BirA protein functions as a biotin protein ligase to regulate biotinylation of an acetyl-CoA subunit. A mutant version of BirA (R118G) can cause promiscuous biotinylation of proximal proteins (248-250). We plan to fuse BirA (R118G) to different TcdA truncations with or without the CROPs, allowing biotinylation of cellular proteins that interact with receptor binding region of TcdA so we could analyze them by mass spectrometry or immunoblot analysis. With this method, one could uncover and confirm potential receptors of TcdA in response to specific binding regions. The BioID method could also be modified as a translocation assay. The BirA protein can be fused to either full length TcdA/TcdB or truncations with GTD removed. Incubation of cells with BirA-TcdB (no GTD) and biotin would allow for biotinylation of human cellular proteins that interact with C. difficile toxins as it internalize and translocate to the cytosol, within minimal cell death during the assay. The difficulty for developing translocation assay for TcdA/TcdB is that the wild type toxins are extremely potent that they cause rapid cell death within short amount of time, making tracking translocation challenging. Even by introducing enzymatically inactive mutants or removing GTD domain, it can still confer cytotoxicity through necrosis. BioID allows us to fuse BirA possibly to fragments of TcdB translocation domain and we can analyze fusion protein stability by BirA activity. Once internalized, the ability of the fusion protein to biotinylate cellular proteins would be a good indication for translocation and also elucidate intracellular routing path for TcdA and TcdB under different conditions (insertion domain, concentration, receptor dependent internalization).

6.4 Concluding Remarks

The overall aim of my thesis was to elucidate the mechanism of cell entry for C. difficile toxins. Both translocation and host cell binding are fundamental steps in toxin intoxication process. Using a combination of electrophysiology, biophysics and cell biology approach, my work uncovered key determinants for TcdB translocation, highlighted the critical role of pore-formation and translocation in cytotoxicity and proposed a double-dagger pore formation model
for TcdB. My discovery of L1106K, a single point mutant in a 2,366 amino acid protein, that was over 1000-fold defective, in addition to helping to understand pore-formation, offered the possibility of developing immunogens for vaccine development (NB a provisional patent for this concept was filed in 2014). Furthermore, our findings revealed the critical region of translocation and receptor binding for TcdB specific receptors and re-established the role of CROP domain in toxin function, addressing some of the pressing issues and knowledge gaps in the field. We anticipate our research on TcdB will help define key principle that underlie translocation and host cell binding which may apply to homologous LCTs.
Appendices

Chapter 2. Translocation domain mutations affecting cellular toxicity identify the *Clostridium difficile* toxin B pore

Appendix Figure 1. Validation of GC-enriched, codon-optimized TcdB protein from *E.coli* expression.

(A) Autoprocessing activity of recombinant toxins. Recombinant TcdB variants were treated with 100 µM InsP6 (+) or PBS (-) for 3h and cleavage was visualized by Western blot by probing with an anti-GTD antibody. (B) GTD activity of recombinant toxins. GST-Rac1 was treated with recombinant toxins, and the level of glucosylation was determined by Western blot analysis using Mab102 that recognizes unglucosylated Rac1 (above) and an anti-Rac1 antibody to determine total Rac1 (below). (C) Functional activity of recombinant toxins. Recombinant TcdB
constructs were added to CHO cells over a range of concentrations. Cellular viability was quantitated 48h later by measuring the fluorescence of cells treated with the cell viability reagent (PrestoBlue®).

Appendix Figure 2. Defective TcdB mutants are folded and display a similar pH-induced unfolding by TNS fluorescence.

(A) Defective mutants were all folded at neutral pH and began unfolding at pH < 5. (B) The two most defective mutants (G1098K and L1106K) show a similar unfolding profile to WT TcdB.
Appendix Figure 3. Alignment of the hydrophobic region of different forms/strains of TcdB.

We aligned five TcdB strains (accession codes): R20291 (YP_003217086), CD196 (YP_003213639), 8864 (CAC19891), 630 (YP_001087135) and 1470 (CAA80815). The eight residues that resulted in a greater than 90% decrease in activity upon mutation are indicate below residues with an arrowhead.
Chapter 4. Crystal Structure of *Clostridium difficile* toxin A

Appendix Figure 4. Wall-eyed stereo view of representative electron density.

Weighted 2Fo-Fc map contoured at 1 sigma.

Appendix Figure 5. A truncated TcdA construct (TcdA<sub>1795</sub>) undergoes autoprocessing more efficiently than TcdA and TcdA<sub>1832</sub>.

The three proteins were incubated with or without 10 mM InsP6 for 2 hours in a buffer containing 20 mM HEPES pH 6.9, 50 mM NaCl and then subjected to SDS-PAGE. Autoprocessing was most efficient in the TcdA<sub>1795</sub> sample. Gel is representative of three independent experiments.
Appendix Figure 6. Zinc is present in the APD active sites of TcdA.

a. Analysis of the anomalous zinc signal reveals a single peak, shown here in purple at 7 sigma, located in the APD active site. The zinc is coordinated by H655 (2.7 Å), C700 (2.7 Å), and H759 (2.2 Å) and may have indirect contacts with E544 or D545. The blue mesh depicts a weighted 2Fo-Fc map contoured around the 5 active site amino acids at 1 sigma. 

b. X-ray absorption near-edge spectrum of TcdA in solution.
Appendix Figure 7. Topology of the TcdA delivery domain. Cylinders indicate α-helices, arrows represent β-strands.
Appendix Figure 8. Full Western blots associated with Figures 3c and 3d.

a. Proteins were separated by SDS-PAGE and transferred to PVDF membrane. The membrane was cut between the 37 kDa and 25 kDa molecular weight markers so that the top blot could be probed with an antibody against the TcdA CROP domain and the middle blot could be probed for unglucosylated Rac1. After exposure to film, the middle blot was stripped and reprobed with an antibody against total Rac1. b. The membrane was initially probed with an antibody that recognizes unglucosylated Rac1. After exposure to film, the blot was stripped and reprobed with an antibody against total Rac1.
Chapter 4. Functional Defects in *Clostridium difficile* TcdB Toxin Uptake Identify CSPG4 Receptor Binding Determinants.

Appendix Figure 9. Domain structures of TcdA and TcdB showing short repeats (green) and long repeats (yellow) in the CROP domains.
Appendix Figure 10. Defective TcdB mutants are folded properly.

**A & B**, Autoprocessing activity of recombinant toxins. Recombinant TcdB variants were treated with 100 µM InsP6 (+) or PBS (−) for 3 h and cleavage was visualized by coomassie staining. Bands were quantified by ImageJ software. **C**, GTD activity of recombinant toxins. GST-Rac1 was treated with recombinant toxins, and the level of glucosylation was determined by Western blot analysis using Mab102 that recognizes unglucosyated Rac1 (top) and an anti-Rac1 antibody to determine total Rac1 (bottom). **D**, pH-induced unfolding by TNS fluorescence. Defective mutants were all folded at neutral pH and began unfolding at pH<5. Both mutants (Y1824K and N1839K) show a similar unfolding profile to WT TcdB.
Appendix Table

Appendix Table 1. Data collection and refinement statistics.

<table>
<thead>
<tr>
<th></th>
<th>Native</th>
<th>HgCl₂ (90 min)</th>
<th>HgCl₂ (3 days)</th>
<th>S1329C HgCl₂ (3 days)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.28149</td>
<td>1.00273</td>
<td>1.00275</td>
<td>1.00798</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>303.61,</td>
<td>300.34,</td>
<td>302.0, 123.99,</td>
<td>300.87,</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>124.54,</td>
<td>75.98</td>
<td>123.80, 75.83</td>
<td>75.76</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.0-3.25 (3.31-3.60)</td>
<td>50.0-3.8 (3.67-4.22)</td>
<td>50.0-3.8 (3.37-4.15)</td>
<td>4.15</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;</td>
<td>6.2 (63.4)</td>
<td>6.9 (50.7)</td>
<td>8.0 (68.2)</td>
<td>8.7 (57.1)</td>
</tr>
<tr>
<td>I/σI</td>
<td>21.0 (2.5)</td>
<td>9.6 (1.8)</td>
<td>10.3 (1.9)</td>
<td>10.4 (2.3)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100 (99.9)</td>
<td>99.3 (99.3)</td>
<td>99.8 (99.8)</td>
<td>99.7 (99.5)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.0 (6.0)</td>
<td>2.7 (2.7)</td>
<td>3.8 (3.8)</td>
<td>5.1 (5.1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>K₂PtCl₃</th>
<th>HAuCl₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>1.06578</td>
<td>1.03501</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>315.88,</td>
<td>297.66,</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>123.41,</td>
<td>75.43</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.0-4.1 (4.17-3.56)</td>
<td>3.5</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;</td>
<td>5.4 (51.9)</td>
<td>6.1 (44.1)</td>
</tr>
<tr>
<td>I/σI</td>
<td>15.9 (2.2)</td>
<td>16.0 (2.5)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (100.0)</td>
<td>99.4 (95.8)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.0 (4.0)</td>
<td>5.2 (4.5)</td>
</tr>
</tbody>
</table>

|                          |          |                |                |
| Refinement               |          |                |                |
| Resolution (Å)           | 50.0-3.26|                |                |
| No. reflections          | 43,813   |                |                |
| R<sub>merge</sub>, R<sub>free</sub> (%) | 18.21/23.74 |                |
| No. atoms                | 14412    |                |                |
| Protein                  | 14411    |                |                |
| Zinc                     | 1        |                |                |
| Water                    | 0        |                |                |
| B-factors (Å<sup>2</sup>) |          |                |                |
| Wilson B                 | 106.15   |                |                |
| Protein                  | 123.95   |                |                |
| Zinc                     | 146.46   |                |                |
| R.m.s deviations         |          |                |                |
| Bond lengths (Å)         | 0.018    |                |                |
| Bond angles (°)          | 1.975    |                |                |

*Highest resolution shell is shown in parenthesis.

<sup>a</sup> R<sub>free</sub> calculated with 4.57% of the reflections (2002) that were omitted from the refinement.
Appendix Table 2. Alignment of enzyme domains to isolated domain structures.

<table>
<thead>
<tr>
<th>Toxin</th>
<th>Ligands</th>
<th>PDB ID</th>
<th>Rmsd Å²</th>
<th>No of Ca atoms</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TcdA</td>
<td>Apo</td>
<td>4DMV</td>
<td>0.65</td>
<td>428</td>
<td>46</td>
</tr>
<tr>
<td>TcdA</td>
<td>Apo</td>
<td>3SS1</td>
<td>0.78</td>
<td>484</td>
<td>24</td>
</tr>
<tr>
<td>TcdA</td>
<td>UDP, Mn</td>
<td>4DMW</td>
<td>0.7</td>
<td>430</td>
<td>46</td>
</tr>
<tr>
<td>TcdA</td>
<td>UDP-glucose, Mn</td>
<td>3SRZ</td>
<td>0.92</td>
<td>455</td>
<td>24</td>
</tr>
<tr>
<td>TcdB</td>
<td>UDP, glucose, Mn</td>
<td>2BVL</td>
<td>1.22</td>
<td>469</td>
<td>47</td>
</tr>
<tr>
<td>TcsL</td>
<td>UDP-glucose, Mn</td>
<td>2VKD</td>
<td>0.96</td>
<td>461</td>
<td>48</td>
</tr>
<tr>
<td>Tcna</td>
<td>Apo</td>
<td>2VK9</td>
<td>1.99</td>
<td>420</td>
<td>48</td>
</tr>
</tbody>
</table>

Appendix Table 3. Zinc to protein ratios determined by ICP-MS

<table>
<thead>
<tr>
<th>Sample*</th>
<th>Zinc/Protein</th>
<th>Sample</th>
<th>Zinc/Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>TcdA&lt;sub&gt;1796&lt;/sub&gt;</td>
<td>0.8, 0.83, 0.72</td>
<td>TcdB</td>
<td>0.74, 0.68, 0.6</td>
</tr>
<tr>
<td>TcdA&lt;sub&gt;1796&lt;/sub&gt; C700A</td>
<td>0.08</td>
<td>TcdB C698A</td>
<td>0.09</td>
</tr>
<tr>
<td>TcdA&lt;sub&gt;1796&lt;/sub&gt; H655A</td>
<td>0.06</td>
<td>TcdB H653A</td>
<td>0.04</td>
</tr>
<tr>
<td>TcdA&lt;sub&gt;1796&lt;/sub&gt; H759A</td>
<td>0.3</td>
<td>TcdB H757A</td>
<td>0.12, 0.31, 0.5</td>
</tr>
<tr>
<td>TcdA&lt;sub&gt;1796&lt;/sub&gt; + InsP6</td>
<td>0.57, 0.51</td>
<td>TcdB + InsP6</td>
<td>0.45, 0.55</td>
</tr>
<tr>
<td>TcdA&lt;sub&gt;1796&lt;/sub&gt; + TPEN</td>
<td>0.04</td>
<td>TcdB + TPEN</td>
<td>0.05</td>
</tr>
<tr>
<td>TcdA&lt;sub&gt;1796&lt;/sub&gt; (apo) + ZnCl₂</td>
<td>0.48</td>
<td>TcdB (apo) + ZnCl₂</td>
<td>0.38</td>
</tr>
</tbody>
</table>

* Each entry represents a measurement from a unique protein preparation.
Appendix Table 4. Primers used to generate the plasmids for this study.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Plasmid</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>TcdA_{1832}S1329C</td>
<td>pBL578</td>
<td>5'-CTTACTCTTTATTATGATGTCATATCCATATCAACG-3' 5'-CTTCTCTTTATTATGATGTCATATCCATATCAACG-3'</td>
</tr>
<tr>
<td>TcdA_{1795}</td>
<td>pBL656</td>
<td>5'-GATAAAATCTTTGATGTCATATCCATATCAACG-3' 5'-GATAAAATCTTTGATGTCATATCCATATCAACG-3'</td>
</tr>
<tr>
<td>TcdA_{1795}C700A</td>
<td>pBL712</td>
<td>5'-GATAAAATCTTTGATGTCATATCCATATCAACG-3' 5'-GATAAAATCTTTGATGTCATATCCATATCAACG-3'</td>
</tr>
<tr>
<td>TcdA_{1795}H655A</td>
<td>pBL713</td>
<td>5'-GATAAAATCTTTGATGTCATATCCATATCAACG-3' 5'-GATAAAATCTTTGATGTCATATCCATATCAACG-3'</td>
</tr>
<tr>
<td>TcdA_{1795}H759A</td>
<td>pBL714</td>
<td>5'-GATAAAATCTTTGATGTCATATCCATATCAACG-3' 5'-GATAAAATCTTTGATGTCATATCCATATCAACG-3'</td>
</tr>
<tr>
<td>TcdB_{H757A}</td>
<td>pBL689</td>
<td>5'-GATAAAATCTTTGATGTCATATCCATATCAACG-3' 5'-GATAAAATCTTTGATGTCATATCCATATCAACG-3'</td>
</tr>
<tr>
<td>TcdASASV1109S, N1110A, N1111S</td>
<td>pBL675</td>
<td>5'-GATAAAATCTTTGATGTCATATCCATATCAACG-3' 5'-GATAAAATCTTTGATGTCATATCCATATCAACG-3'</td>
</tr>
</tbody>
</table>
References


76. Weiss, K. (2009) Toxin-binding treatment for Clostridium difficile: a review including reports of studies with tolevamer. *International journal of antimicrobial agents* 33, 4-7


Mechanisms of protection against Clostridium difficile infection by the monoclonal antitoxin antibodies actoxumab and bezlotoxumab. *Infect Immun* **83**, 822-831


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


133. Cherfils, J., and Zeghouf, M. (2013) Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiol Rev* 93, 269-309


149. May, M., Wang, T., Muller, M., and Genth, H. (2013) 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 5, 106-119

150. Hecht, G., Koutsouris, A., Pothoulakis, C., LaMont, J. T., and Madara, J. L. (1992) Clostridium difficile toxin B disrupts the barrier function of T84 monolayers. Gastroenterology 102, 416-423


169. Daiber, A. (2010) Redox signaling (cross-talk) from and to mitochondria involves mitochondrial pores and reactive oxygen species. *Biochim Biophys Acta* 1797, 897-906


144


transmembrane domain is essential for the efficient delivery of the catalytic domain to the cytosol of target cells. *J Biol Chem* **268**, 12077-12082


