STRUCTURAL ANALYSIS OF *CLOSTRIDIUM DIFFICILE* TCDB UPTAKE MECHANISM

A Dissertation

by

MENGQIU JIANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Hays Rye
Committee Members,	Junjie Zhang
	Michael Manson
	Jason Gill
Head of Department,	Joshua Wand

May 2021

Major Subject: Biochemistry

ABSTRACT

Clostridium difficile is a gram-positive, anaerobic, spore-forming bacteria that causes severe diarrhea, abdominal pain, and pseudomembranous colitis (1). *Clostridium difficile* infection (CDI) can be fatal. The estimated annual cost for CDI management is around $6.1 billion in the U.S.. When treating CDI using broad-spectrum antibiotics, the rate for CDI recurrence increases significantly. Thus, developing new treatments for CDI is essential. TcdB is the major virulence factor secreted by *Clostridium difficile,* and it is responsible for most of the symptoms of CDI. It is critical to understand its mechanism of entry into intestine epithelial cells and subepithelial layers to devise methods to block as many of the uptake steps as possible. In this dissertation, I have studied the interaction of TcdB with its two receptors, frizzled-2 (FZD2) and chondroitin sulfate proteoglycan 4 (CSPG4), using cryo-EM and biochemical assays. TcdB utilizes hydrophobic interactions to bind FZD2 and CSPG4, and CSPG4 also has electric charge interactions with TcdB. Three snapshots of TcdB binding to CSPG4 were obtained, demonstrating that TcdB binds in an equilibrium among different states. I also analyzed TcdB binding at acidic pH and found that there is a large conformational change in the overall structure, primarily in the hydrophobic region of the delivery domain. Both of the receptors dissociated with TcdB when pH decreased, with CSPG4 dissociation at a larger rate, allowing TcdB to float near the endosome membrane for the following pore formation and translocation. By revealing detailed receptor binding mechanism as well as intermediate states of TcdB when the decrease of the pH triggers the conformational

changes, this work extensively expand a structural view of TcdB uptake mechanism and provides strong reference to resolve homolog toxin invasion mechanism. To develop a potential treatment for CDI, a library of Designable Ankyrin Repeat Proteins (DARPins) was generated to screen for TcdB neutralizer. A dimeric DARPin, DLD-4, that is composed of two monomeric DARPins, U3 and 1.4E, is selected because it has the best neutralization ability as well as high binding affinity to TcdB. My cryo-EM structure of TcdB-DLD4 complex demonstrates that U3 interacts at the FZD2 receptor binding surface and 1.4E grips on the CSPG4 interaction region. Thus, DLD-4 is able to neutralize the TcdB by competing against its receptors with much higher binding affinity and block TcdB entry at the first step.

DEDICATION

I dedicate this work to my beloved husband Yaxi Wang and my son Russell Wang for their support.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Junjie Zhang during my Ph.D period. He taught me how to be a dedicated researcher and trusted me and supported me when I was struggling with my studies. He also encouraged me and provides me solid background to finish my work.

I would also thank my committee chair, Hays Rye. He is always patient to my questions and support me on the experiments that I need help with. For example, the leakage assay on TcdB and liposome was conducted under his supervision. my committee members, Dr. Manson and Dr. Gill, for their guidance and support throughout the course of this research.

My colleagues in the lab also provide me a lot of help. Karl Gorzelnik is knowledgeable in many experiments and helped me on learning new skills. Zhicheng Cui and Jeng-yih Chang helped me with data analysis and the computational tricks. Ran Meng and I worked together on the earlier data collection of TcdB and FZD2 manually. Jirapat Thongchol is willing to help whenever I need.

My collaborators are always willing to help me to conduct the experiments that I am not able to do and we communicated actively. Their efforts make this happens. Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Junjie Zhang and Professor Hays Rye of the Department of Biochemistry and biophysics, and Professor Michael Manson of the Department of Biology and Professor Jason Gill of the Department of Animal Science.

All the ELISA assays and DARPin screens in the main text were conducted by Dr. Rudo Simeon from Dr. Zhilei Chen's laboratory and the content in chapter IV were published in 2019. The leakage assay was performed by Xue Gong from Dr. Hays Rye's laboratory. The biochemical analysis in Appendix A was done by Jeremy Weaver and published on *Nature Communications* at 2017. The cryo-EM data collection was done in SLAC cryo-EM facility in Stanford.

Funding Sources

This work was made possible in part by startup funding from the Department of Biochemistry and Biophysics at Texas A&M University and Center for Phage Technology jointly sponsored by Texas AgriLife, Texas A&M University and Welch Foundation grant A-1863 and NIH grants R21AI137696 and P01AI095208.

.

NOMENCLATURE

GI tract	Gastrointestinal tract
GTD	glucosyltransferase domain
APD	Autoprocessing domain
CROPS	Combined Repetitive Oligopeptides
FZD2	frizzled family protein-2
CSPG4	chondroitin sulfate proteoglycan 4
MC	multiple cloning sites
PCR	Polymerase Chain Reaction
E.coli	*Escherichia coli*
LB	Luria-Bertani
PBS	Phosphate-buffered saline
ELISA	enzyme-linked immunosorbent assay
IPTG	Isopropyl β- d-1-thiogalactopyranoside
SDS-PAGE	sodium dodecyl sulphate–polyacrylamide gel electrophoresis
CTF	Contrast Transfer Function
DARPin	designed ankyrin repeat proteins
FSC	Fourier Shell Correlation
CD	circular dichroism
DDM	n-Dodecyl β-D-maltoside
DT	diphtheria toxin

C. difficile	*Clostridium difficile*
LCT	Large Clostridial Toxin

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

1.1. Overview

Clostridium difficile (C. difficile) was first isolated from a healthy infant in 1935 (2) and identified as a spore-forming, anaerobic bacterium of the Bacillus family (3). It was given the species name *difficile* for the reason that it is hard to isolate because of its slow growth rate. As it could be isolated from healthy adults and children, *C. difficile* was not considered to be detrimental in the human digestive tract. Because of its slow growth rate, it always remains a minor component of a healthy human digestive tract, which is protected by a group of bacteria collectively called "healthy flora." However, after administration of broad-spectrum antibiotics, such as lincomycin and vancomycin (4) (5), this protective flora is wiped out, allowing *C. difficile* spores to settle and begin to germinate in the small intestine, producing toxins in the large intestine that destroy the colonic epithelium tissue (6). The damage increases the permeability of the epithelial barrier, allowing the toxins to permeate to deeper colonic tissues.

The two Large Clostridial Toxins (LCT), enterotoxin TcdA and cytotoxin TcdB, as well as the binary toxin CDT, are responsible for nearly all the symptoms caused by *C. difficile* infection (4). There are several factors that lead to *C. difficile* infection. Hospitalization is one of the factors as the *C. difficile* spore is hard to be eliminated and enriched on the surfaces of the hospital equipment. These spores are able to be subsequentially transmitted through fecal-oral pathway. Exposure to antibiotics disrupts the homeostasis of microflora in the colon that provides opportunity for *C. difficile*

spores. Ages also influence the chances for *C. difficile* germination as the old or very young age has fairly weak immune system. (7). In 2007, CDI was the leading cause of gastroenteritis-associated death (8). It was considered to be the most common cause of hospital-acquired infection in the United States, and the estimated health care cost related to CDI in acute care facilities was around $4.8 billion (9). An additional problem is that, because *C. difficile* is a spore-forming bacterium, it often causes secondary infections, known as recurrence, when antibiotics are applied to combat the initial treatment (10). Therefore, treatments for CDI must be chosen carefully, and alternatives to antibiotics should be developed to relieve the colitis symptoms while avoiding a high rate of recurrence.

1.2. *Clostridium difficile* biology

Clostridium difficile is a spore-forming, gram-positive anaerobic bacillus that can colonize the human gastrointestinal tract (GI tract) (11) (12). The survival of *C. difficile* involves in two important steps: sporulation and germination. The ability of *C. difficile* to sporulate is critical both for the initial onset of CDI as well as its recurrence (13). Germinate in the appropriate place makes it possible for *C. difficile* to reproduce and generate offspring. Thus, understanding both sporulation and spore germination is important for developing treatments for CDI.

1.2.1. Life cycle of *Clostridium difficile*

C. difficile has two forms: the spore, which is resistant to harsh environments, and a vegetative form that germinates and proliferates in the host's digestive tract (14). The vegetative form also produces the toxins that cause symptoms that vary from mild

diarrhea to severe pseudomembranous colitis, and even death (15). As vegetative *C. difficile* is strictly anaerobic, the spore form of *C. difficile* is required for its survival outside the host and for its transmission. The transmission of *C. difficile* is through fecal-oral pathway (16). Only after the dormant *C. difficile* spore enters the host's digestive tract can germination begin. When the spore gets into the small intestine, the bile salts in the environment serve as signals, called germinants, that trigger germination (17). As germination proceeds, *C. difficile* continues to move through the digestive tract into large intestine and produces vegetative cells (13), which reproduce and secret toxins. The toxins permeabilize the epithelial tissue, leading to diarrhea. Simultaneously, a small portion of vegetative cells transform into spores that are eliminated with the patient's stool, ready to infect the next patient (Figure 1-1).

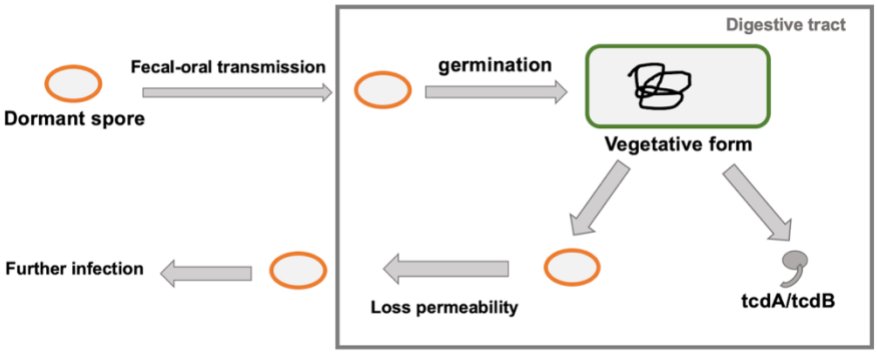


Figure 1-1 Scheme of *C. difficile* life cycle.

1.2.1.1. Sporulation of *C. difficile*

The *C. difficile* spore is critical for survival outside the host, and therefore also for transmission. 1). *C. difficile is* a strictly anaerobic bacillus, and the spore protects its genome and other essential cellular machinery from damage in the environment; 2). The dormant spore has nearly no cellular activities, making it intrinsically resistant to antibiotic treatment and the bleach-free disinfectants that are usually used in the hospital. As its center is wrapped by multiple condensed layers of peptidoglycan, *C. difficile* spore is resistant to any harsh environment, such as extreme pH, heat, and radiation.

C. difficile initiates sporulation in response to a stimulus, such as the limitation of nutrients, specific signal chemicals, quorum sensing, radiation, and other stress factors (18) (19). As in other bacilli, *C. difficile* sporulation involves four steps (20) (21). The first step is the asymmetric division into a mother cell (MC) and a smaller compartment (SC). The SC will then gradually form a fore spore that is completely encapsulated in the MC. The third step is the assembly of the cortex and coat layers during maturation of the spore. Finally, the mature spore is released into the environment after the MC lyses (20). The structure of the *C. difficile* spore is similar to that of the spores of other endospore-forming bacteria. The core contains genomic DNA, mRNA, ribosomes, proteins and high concentrations of pyridine-2,6-dicarboxylic acid (DPA) to protect the genomic DNA from heat damage (19). The *C. difficile* DNA is supercoiled and bound with small, acid-soluble proteins (SASPs), blocking the transcription and protect DNA from damage. The spore core is enclosed within an inner membrane, with additional layers

made up of a peptidoglycan cortex, an outer membrane, and coat proteins. These layers protected the spore from damage by environmental stress (Figure 1-2).

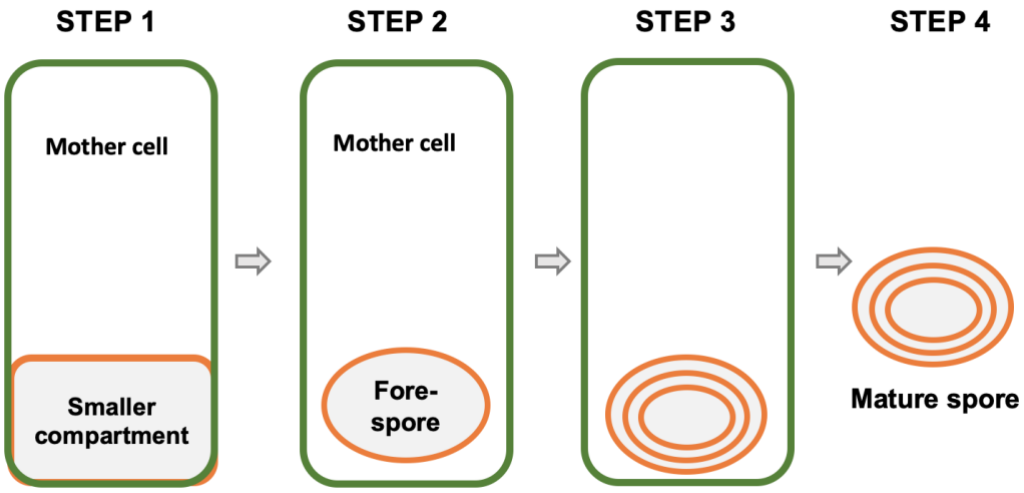


Figure 1-2 Scheme of *C. difficile* sporulation.

The sporulation of *C. difficile* is negatively regulated by CodY and CcpA, which are nutrient sensors (22). The detailed mechanism of CodY regulation is not fully understood, but it negatively regulates the expression of sporulation-related genes, such as s*po0A, rapA, and rapC*. CcpA is a carbohydrate-sensing protein that regulates overall gene expression. CcpA directly regulates the *ParLoc* genes, including *tcdC, tcdR, tcdA and tcdB*, through glucose-dependent repression (23) (24).

Like many other bacilli, the sporulation of *C. difficile* is controlled by the master transcriptional regulator Spo0A, which is a DNA-binding protein that regulates entry into the sporulation pathway. Without Spo0A, all sporulation activity is lost. The activity of Spo0A is regulated by its phosphorylation by an orphan histidine kinase. Even though *C. difficile* does not have the same histidine kinase (*Kin*) as *Bacillus,* five putative orphan kinases {CD1352 [CD630_13520; cprK (McBride and Sonenshein, 2011)], CD1492 (CD630_14920), CD1579 (CD630_15790), CD1949 (CD630_19490), and CD2492 (CD630_24920)} that phosphorylate Spo0A were identified in *C. difficile* strain 630 (25).

After Spo0A is phosphorylated, it is able to regulate sporulation through the sporulation-specific RNA polymerase sigma factors σ^F, σ^E , σ^G , σ^K, which are involved in different sporulation steps (26). For example, σ^K functions downstream of σ^E to regulate the late stages of sporulation. There are some differences in the properties of the *C. difficile* sigma factors compared to *Bacillus*. For example, in *C. difficile* the activation of σ^G does not require σ^E, and the activation of σ^K doesn't require σ^G (27).

1.2.1.2. Germination of *C. difficile*

When the *C. difficile* spore is ingested in the patients intestine, transition to the vegetative form is critical for the following outgrowth and the secretion of toxins in the large intestine. Thus, the germination of the *C. difficile* spore is an essential step prior to its outgrowth and secretion of diarrhea-related toxins. The germination of *C. difficile* involves three steps. Germination is initiated by the germinant in the host environment binding to its receptor in the inner spore membrane (28). Unlike *Bacillus*, *C. difficile*

uses bile salts and some amino acids, such as glycine, as major germinants, since bile salts are commonly found in the GI tract (29). Upon receiving these signals, large amounts of cations as well as Calcium- dipicolinic acid (Ca-DPA) molecules are released from the spore core in exchange for water from the environment. This hydration activates the lytic enzymes in the spore cortex, which degrade the peptidoglycan cortex layer, leading to rehydration and resumption of metabolism in the spore core (20) (21).

The germination of many spore-forming bacterium is induced by the presents of specific small molecules named germinants, indicating a suitable environment for colonization. The germinants are usually nutrients such as sugar, amino acids or nucleotides. *gerA* family genes encode the germinant receptors that locate on the spore inner membrane and trigger the germination initiation. However, there is no *gerA* genes existing in the *C. difficile* genome, indicating that *C. difficile* utilize a unique mechanism of germination initiation (30).

In addition to sensing the nutrient germinants, *C. difficile* spore germination is also in response to bile salts existence. In small intestine, there are two sets of bile salts related to *C. difficile* gemination: primary bile salts such as cholate and taurocholate and secondary bile salt such as deoxycholate and chenodeoxycholate. Most of the primary bile salts are absorbed in small intestine and transferred to the liver, forming into conjugated bile salts. The remained primary bile salt will then loss a hydroxy group and turned into secondary bile salts with the present of flora (31)(Figure 1-3). It is showed that primary bile salts are essential germinants for *C. difficile* spore, in combination with the presents of amino acids or calcium ion, while the secondary bile salts inhibit *C.*

difficile germination. The receptor for primary bile salts in *C. difficile* is a pseudoprotease CspC. This pseudoprotease is defined as catalytically inactive enzymes. The binding of primary bile salts to CspC synergically enhanced the binding of cogerminant glycine or calcium. However, the receptor of the cogerminants remains unknown (32).

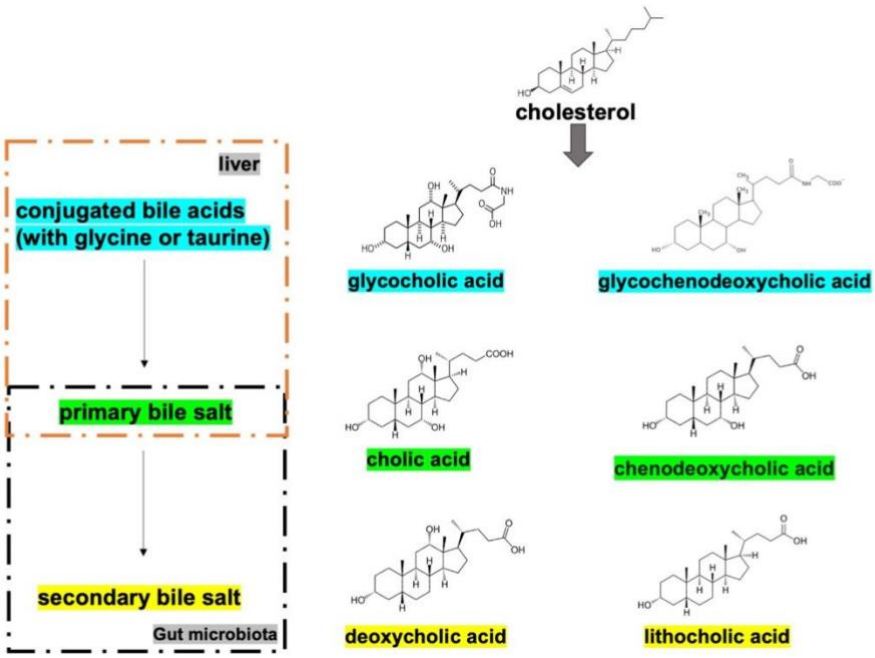


Figure 1-3 Examples of bile salts
The primary bile acids are derived from cholesterol and they usually conjugate with glycine or taurine in the liver. When secreted in the small intestine, they exist as primary bile acids, which are dehydroxylated by gut micro-bacterium into secondary bile acids.

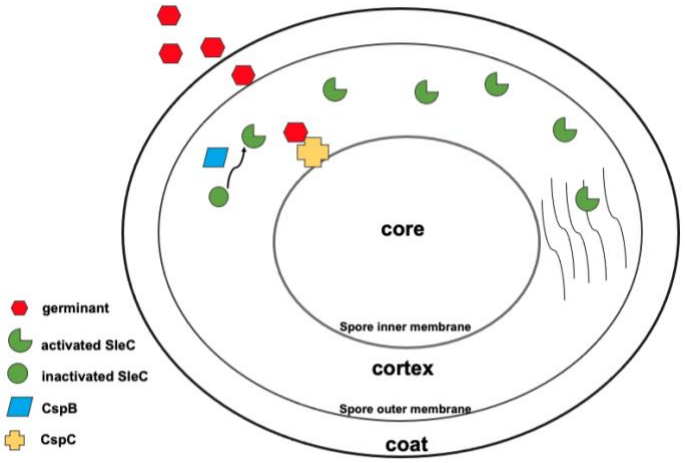


Figure 1-4 The scheme of the *C. difficile* germination activation
The germinants get in touch with the CspC receptor located on the spore inner membrane. This interaction activates the CspB to cleave the C-terminus of SleC, releasing active form SleC that is able to digest the peptidoglycan in the cortex.

After binding with the germinant and activating the germination process, the cortex lytic enzyme SleC is sequentially activated by cleavage of prodomain from SleC by Csp family of subtilisin-like proteases (CspABC). In *C. difficile*, there is a gene operon *cspBAC* encodes fusion protein CspBA, which will be cleaved into CspB and CspA proteins before wrapped into mature spore. CspB is required for germination as it directly cleaves the prodomain from SleC protein. CspA helps the incorporation of CspC into mature spore. As the cortex of *C. difficile* spore is hyper-dense and remains high salt concentration, the hydrolysis of cortex results in hypo-osmotic stress in the core of C. difficile spore. Releasing Ca-DPA will relieve this stress in exchange of water molecule from the environment (Figure 1-4). Thus, the *C. difficile* spore retain its ability to the cellular activities and transformed into vegetative form *C. difficile* (33).

1.2.2. Pathogenesis

To develop CDI, an individual needs to meet these two prerequisites: 1). The disruption of homeostasis of microflora in intestine. The diminished symbionts lost their resistance to *C. difficile* spore and provide chances for its to germinate. 2). The individual needs to acquire *C. difficile* spore from exogenous source through fecal-oral transmission. Nearly all antibiotics that eliminate the resident bacteria are able to cause infection by *C. difficile*, but utilization of broad-spectrum antibiotics is the most common cause. After antibiotic treatment, *C. difficile* can overgrow all the other enteric bacteria and form heat-resistant spores, which can persist for months or years (34).

Much of the equipment in hospitals is contaminated with *C. difficile* spores (35). Another source of *C. difficile* spores is from other patients that have diarrhea. After colonization, pathogenic *C. difficile* produces toxins that cause diarrhea and pseudomembranous colitis. Even after recovery from a *C. difficile* infection, a patient might be exposed again and get a recurrence of CDI (36). The basis for the recurrence of CDI is still not well understood. Based on the current clinical data, the first recurrence of CDI occurs with a frequency of approximately 20%, with subsequent recurrence (37). Recurrence depends on factors such as the antibiotics used and whether the conditions for spore germination exist (38).

1.2.2.1. Toxins secreted from *C. difficile* leads to CDI symptoms

Toxins are secreted by the vegetative form *C. difficile* primarily in the large intestine (39). Two gene loci are related to the secreted *C. difficile* toxins: the pathogenicity locus *(PaLoc)* that encodes two large single-chain toxins, and *CdtLoc* that

produces binary toxin (40) (41). *PaLoc* is a 19.6kb region that encodes TcdD, TcdE, TcdR, TcdA, and TcdB. The binary toxins CDTa and CDTb are encoded by *CdtLoc* (42). This locus also contains the regulatory gene *cdtR* (43). All of these toxins are able to disrupt the assembly or disassembly of actin, resulting in cell rounding and an increase in the permeability in the epithelial tissue (44).

TcdA and TcdB modify the small GTPase in the colonic tissues, causing loss of tight junctions in the epithelium and cell death. They are responsible for most of the symptoms associated with CDI (45). TcdD, TcdE, and TcdR regulate the expression and secretion of TcdA and TcdB (46) (47). While most ribotypes of *C. difficile* expresses both TcdA and TcdB, strains that express only TcdA or TcdB exist, and they are also infectious (48). As shown in Figure 1-5, *tcdC* is transcribed in the opposite direction from *tcdA* and *tcdB* (49). It is highly expressed in early exponential phase, and the expression level of TcdC significantly decreases when *C. difficile* reaches stationary phase (50). This decrease corresponds to increased expression of TcdA and TcdB, suggesting that TcdC is a repressor of TcdA and TcdB expression.

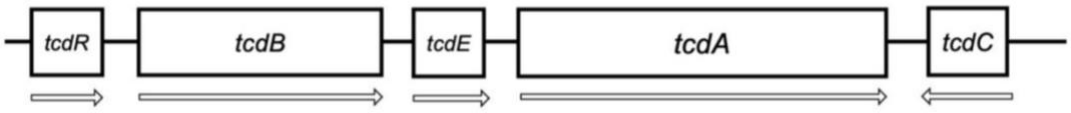


Figure 1-5 Gene organization of *PaLoc*
The genes in *PaLoc* are represented by boxes, with their name listed. The transcription directions of each gene are indicated by arrows.

TcdD is predicted to be a DNA-binding protein from its structure (51). It has been demonstrated that TcdD is an alternative sigma factor for *C. difficile* RNA polymerase (52). The TcdD RNA polymerase holoenzyme transcribes the downstream genes *tcdA* and *tcdB*. Expression of TcdD is significantly elevated in the stationary phase (50), during which *C. difficile* secretes large amounts of TcdA and TcdB. TcdE has been identified as a homolog of phage holin proteins, and it is speculated to create holes in the *C. difficile* membrane that are required for the secretion of TcdA and TcdB (53) . The detailed mechanism of TcdE function remains to be understood. The production of TcdA and TcdB depends on various factors such as the temperature, the presence of glucose, certain amino acids, and antibiotics (24).

TcdA and TcdB are both characterized as large clostridial toxins (LCTs), which are expressed by bacteria in the clostridium family. Both TcdA and TcdB disrupt actin assembly by modifying small GTPases in the cytoplasm. This modification irreversibly adds glucose from UDP-glucose to Rho, Rac, and Cdc42, which disrupts many vital signaling pathways (54), including maintenance of the cytoskeleton. The results is cell rounding and increased permeability of the intestinal epithelium.

The molecular mechanism of the action of TcdA and TcdB have been studied for decades. TcdA and TcdB are very similar in their amino acid sequence and their protein domain composition, and they work by similar mechanisms (55). TcdA is the largest LCT, with a molecular weight of 305kD, which is ~50kD bigger than TcdB. TcdA and TcdB have 41% sequence identity and have the same domain organization. They have an N-terminal glucosyltransferase domain (GTD), followed by an autoprocessing domain

(APD), a delivery domain, and a compact repetitive oligo peptides (CROPS) domain. The extra 50kD of TcdA relative to TcdB consists of a longer CROPS domain. All of the other domains are very similar. The GTD domain is the enzyme that modifies the GTPase in the host cell (56). The APD domain cleaves and releases the GTD domain into the cytosol (57). The delivery domain serves as a scaffold for the toxin and contains a hydrophobic region that is responsible for delivering the GTD and APD domains from the inside to the outside of the endosome (58). For TcdA, the CROPS domain is identified as a receptor binding region that allows the entry of TcdA into the cell (59). TcdB gets into the host cell through different mechanism. Three membrane proteins are proposed as TcdB receptors. Through biochemical analysis, they might interact with TcdB in different domains. For example, Frizzled-2 protein, that is involved in the wnt signaling pathway, interacts with TcdB at the delivery domain. The N-terminus of CROPS domain might also be involved in the conformational change in the toxins that occurs when the pH in the endosome decreases as the early endosome turns into late endosome (60).

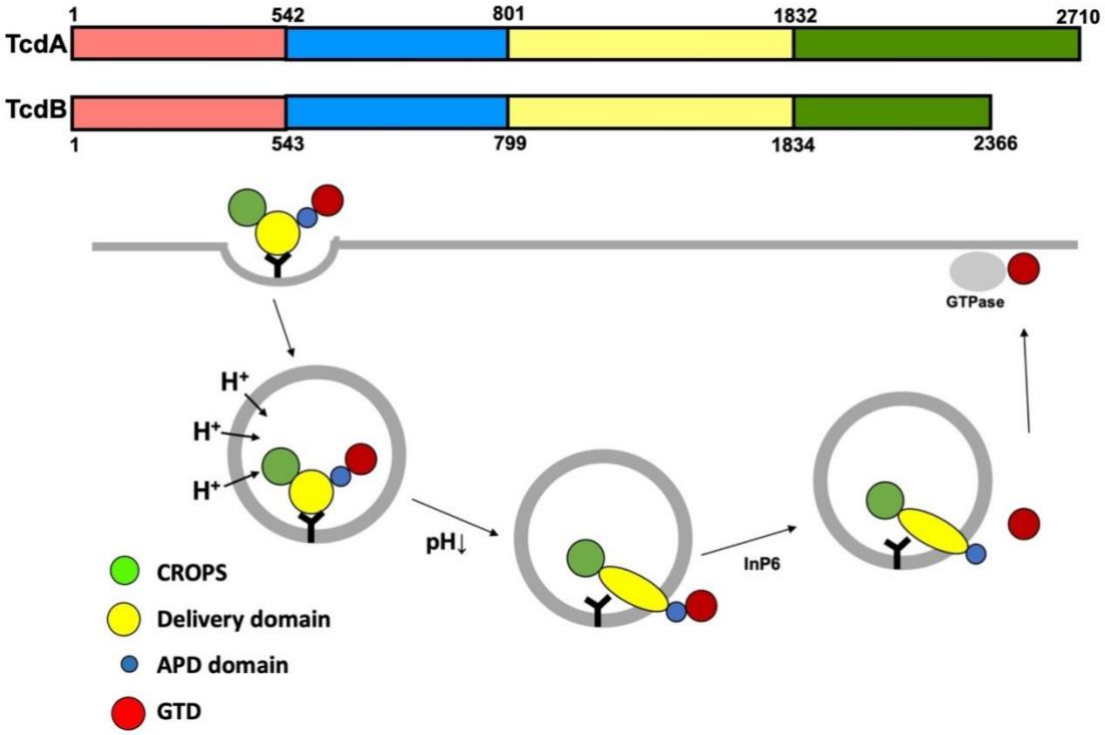


Figure 1-6 Structure organization of TcdA and TcdB, and scheme of TcdB uptake mechanism.
The domain organization of TcdA and TcdB are showed on the top, with designated color for each domain. The bottom panel shows the scheme of TcdB uptake process.

In order to enter the epithelial cell, both toxins need to bind to their receptors on the cell membrane. The uptake involves the endocytic pathway. Toxins are enclosed in the early endosome, whose lumenal pH decreases as the early endosome turns into the late endosome. This pH change triggers a conformational change in the toxins, exposing a hydrophobic region in the delivery domain that allows the toxin to insert into the endosome membrane (61). As a result, the GTD and APD domains protrude into the

cytosol. In the presence of inositol hexaphosphate, the APD domain is able to self-cut and release the 60kD GTD domain (62) (Figure 1-6). The released GTD domain requires the TRiC chaperone in the cytosol for proper folding into its enzymatically active form that glycosylates the small GTPases, thereby disrupting normal actin assembly and disassembly and destroying the cytoskeleton (63).

1.2.2.2. Disruption of host cell physiology by TcdA/TcdB uptake

As mentioned above, the uptake of TcdA/TcdB leads to disruption of the cell skeleton. As different kinds of GTPase regulate the polymerization of actin, toxicity from TcdA/TcdB results in morphological changes in the host cells. Cells that have taken up TcdA/TcdB display different retraction phenotypes than the normal cell. After 2hr of infection, host cells exhibit a rounding phenotype, and cell death ensues one day infection, suggesting that it is not the loss of cell morphology per se that is lethal (64). In addition to disrupting cell morphology, apoptosis can be triggered by TcdA/TcdB when GTPase-related signaling pathways are disrupted (65). For example, Cdc42 and Rac are important for regulation of the cell cycle because of their role in signaling through mitogen-activated protein kinase kinases (MAPKKs) (66). Once bound with GTP, small GTPases such as Ras will activate Raf kinase. This starts a cascade of phosphorylation in the MAPK signaling pathway in response to an external stimulus (67). Once the small GTPase is paralyzed, the host cell is no longer able to react to environmental changes or regulation by the human body. TcdA/TcdB also increase the permeability of the cell by disrupting the function of the RhoA protein, resulting in malfunction of the downstream protein kinase C. Dysfunction of Rho also activates Caspase-3 and Caspase-9, leading to

apoptosis of the host cell (68). The sum of the effects of TcdA/TcdB infection leads to cell death, even at a fairly low dose compared to what is required with the other LCT toxins.

1.3. The mechanism of TcdB enter and affect intestine epithelial cell

1.3.1. Overview of TcdA and TcdB infection

TcdA and TcdB intoxicate host cells through a multistep mechanism: 1) TcdA and TcdB enter the cell through the endocytic pathway and are localized in the early endosome; 2) The ATPase in the early endosome membrane pumps H^+ into endosome so that the pH of the lumen constantly decreases. The acidic pH induces large conformational changes in the pore-forming region of delivery domain of TcdA or TcdB. These conformational changes result in the delivery of GTD and APD from the lumen of the endosome into the cytosol. 3) In the presence inositol hexaphosphate in the cytosol, GTD is released from other domains by the autoprotease activity of the APD domain. 4) The released GTD then uses UDP-glucose as substrate to add mono-glucose onto target GTPases, which leads to a disruption of cellular activities. Although this general outline is well established, the details of the interaction between TcdA/TcdB and their receptors are not known, and the conformational change in the delivery domain is not clear.

1.3.1.1. TcdB receptor binding

Receptor binding is the first step of TcdB entry into the cell and is critical in the development of CDI. Even though TcdB and TcdA share 41% protein sequence identity (55), they probably utilize different receptors for uptake. The receptors for TcdA and

TcdB have been studied for decades, and several have been identified. The C-terminal CROPS domain was initially identified as the receptor-binding region for TcdA and TcdB. Several years later, it was shown that the TcdB CROPS domain is involved in receptor binding through a different mechanism against TcdA CROPS domain (69). As TcdA and TcdB are very similar in sequence, their CROPS domains are predicted to adopt a similar structural organization. The CROPS domain consists of multiple short repeats (SRs), each of 19-24 amino acids, and long repeats (LRs) containing 31 amino acids (59) (70). The TcdA CROPS domain contains 33 SRs and 7 LRs. TcdB has a much shorter CROPS domain that contains 21 SRs and 4 LRs (71). The crystal structures reveal that the SRs and LRs form beta-sheets. The beta-sheets of the SRs stack up against each other with a 120-degree angle, resembling the structure of a solenoid, a similar structure like the assembly of supercoiled DNA in eukaryotic cells. When an SR interacts with a LR, they form a 90-degree screw-axe transformation (72). This structure increases the interaction surface of the protein that can interact with other proteins or saccharides (Figure 1-7).

TcdA was found to interact with carbohydrates on mammalian cell membrane glycoconjugates. The trisaccharide Galα1–3Galβ1–4GlcNAc was shown to interact with TcdA and was proposed to be an important receptor for TcdA. The carbohydrate-binding groove is formed by an LR and the following SR (73) (74). Since TcdA CROPS domain consists of 7 LR and 33SR, it has seven binding pockets to receptor carbohydrate throughout the CROPS domain. One piece of evidence is that when TcdA is occupied by a CROPS-domain-specific antibody, it cannot be taken up into host cells. Sucrase-

isomaltase (SI), a glycosylated protein, is proposed to be one of the TcdA receptors (75). Treating the SI protein with alpha-glucosidase results in loss of interaction between TcdA and SI, supporting the idea that TcdA interacts with SI through its covalent-linked carbohydrate. Glycoprotein gp96, expressed in colonocytes, has also been identified as a TcdA receptor (76). Although it has been speculated that TcdA binds to the N-linked glycan of gp96, there is no evidence to support that TcdA interacts directly with the gp96 glycosylation site, and the residues of TcdA that are involved in this interaction are not conserved in other clostridial glucosylating toxins (77).

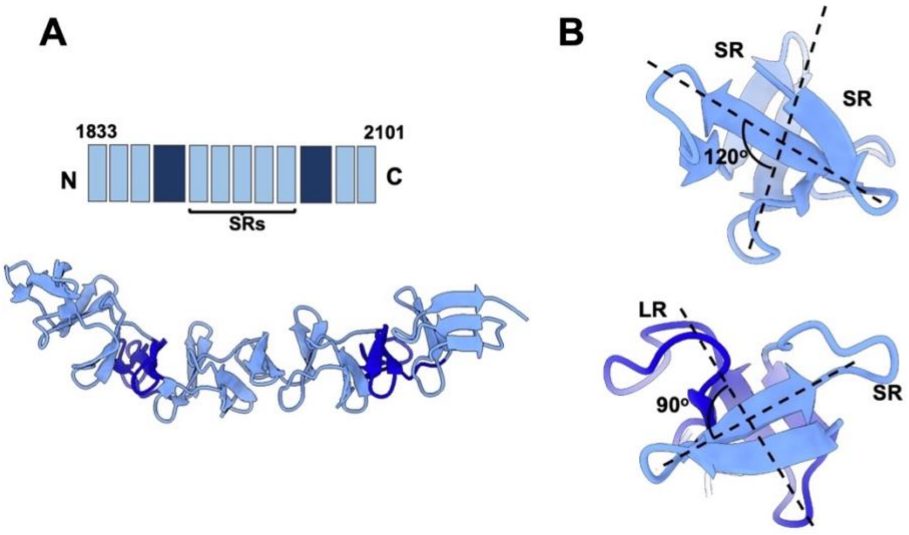


Figure 1-7 Structure organization of TcdB CROPS domain.
The scheme of a CROPS domain fragment is showed in panel A, with short repeat (SR) colored in light blue and the long repeat (LR) colored in deep blue. Panel B shows the angles formed by two SRs at 120º or formed by one LR and one SR at 90º.

Interestingly, TcdB was not found to interact with membrane saccharides. It is highly possible that TcdB has a receptor interaction mechanism that is totally different from that of TcdA. From a whole-genome human shRNAmir library screen, one TcdB receptor was identified. It is the membrane protein chondroitin sulfate proteoglycan 4 (CSPG4) (78). CSPG4 was first identified interacting with the TcdB CROPS domain. By doing the interaction test with different truncated variants of TcdB, it was concluded that CSPG4 interacts with the 1831-1851 region, which is at the N-terminus of the CROPS domain (69). However, it is noteworthy that cell lines that do not express CSPG4 can still be sensitive to TcdB toxicity, suggesting that other TcdB receptors remained to be identified.

Several years later, two additional potential receptors of TcdB were identified as poliovirus receptor-related 3 (PVRL3) (79) and frizzled family proteins, especially frizzled-2 (FZD2) and frizzled-7 (FZD7) (80). Both of these protein types were demonstrated to interact with TcdB in non-CROPS domain. Through genome sequencing using TcdB-challenged Hela cells, FZD1/2/7 were identified as potential receptors for TcdB entry (80). A recently published study solved the crystal structure of a complex of a TcdB fragment and the FZD2 extracellular domain. Central to the binding was a hydrophobic interaction between the middle region of the TcdB delivery domain and the FZD2 extracellular domain. A palmitoleic acid (PAM) inserted at the interaction surface significantly enhances the binding affinity. The interaction of TcdB and FZD2 also obstructs the downstream Wnt signaling pathway (1).

PVRL3 belongs to the protein family PVRL1 to PVRL4. These proteins have three extracellular Ig-like domains. Purified PVRL3 has been reported to interact with purified TcdB, and an anti-PVRL3 antibody decreased TcdB toxicity. PVRL3 and FZD2 are both expressed on epithelial cells of the intestine, and TcdB is able to be taken up into these cell and destroy the tight junctions within the epithelial tissue (79). CSPG4 is expressed in subepithelial myofibroblasts and can serve as a TcdB receptor after the surface epithelium is ruined. This will lead to further damage on the patient's colon. TcdB seems to interact with CSPG4 and FZD2 in an independent and additive manner, meaning that cells that express both receptors are more susceptible to TcdB toxicity (81).

1.3.1.2. TcdB pore forming and translocation in the late endosome

TcdB enters the host cell through the endocytic pathway after binding to its receptor(s). During this process, clathrin accumulates under the membrane where TcdB interacts with the receptor, wrapping TcdB into the early endosome. The early endosome develops into the late endosome when the endosomal ATPase pumps H^+ into the endosome lumen. With the entry of H^+, the pH in lumen decreases, inducing conformational changes in TcdB. When the H^+-ATPase activity is blocked, host cells are no longer susceptible to TcdB toxicity, supporting the idea that reduction of endosomal pH is essential for TcdB toxicity (82). Diphtheria toxin from *Corynebacterium diphtheriae* and anthrax toxin from *Bacillus anthracis* also adopt significant conformational changes upon pH reduction, leading to exposure of a specific hydrophobic region, which will subsequently insert in the host cell membrane and create holes on the endosome (83) (84). Because TcdB is another type of AB toxin, it is

reasonable to speculate that TcdB goes through similar structural changes to create holes.

This prediction was confirmed by directly observing the conformational changes of TcdB. When TcdB was labeled with the fluorescent probe 2-(p-toluidinyl) naphthalene-6-sulfonic acid (TNS), an increase in the fluorescent signal was detected when the pH was lowered to 4.5, suggesting that TcdB undergoes conformational changes upon acidification (85). A leakage assay was performed with CHO cells preloaded with Rb^+(86)and treated with TcdB. When the pH reached 4.5, there was a significant release of Rb^+ into the supernatant, suggesting that TcdB creates holes in the cell membrane (86). This phenomenon was confirmed using artificial lipid bilayer membranes. TcdA is also reported to have similar properties, although TcdA requires cholesterol to present in order to create holes in the membrane. Cholesterol does not influence hole formation by TcdB (87).

Like other toxins that are susceptible to pH changes, TcdB has a hydrophobic region that becomes exposed at low pH, and TcdB changes its conformation from a pre-pore to a pore-forming state. This hydrophobic region lies in the middle of the delivery domain. The crystal structure of $TcdA_{4\text{-}1802}$ revealed a unique delivery domain, with an elongated rod shape extending from the core region formed by GTD and APD (88). This rod has a scaffold of multiple $\boldsymbol{\beta}$-sheets, with four hydrophobic $\boldsymbol{\alpha}$-helix, linked by short loops, wrapping around the surface of the delivery domain (88). A recently published paper reported a crystal structure of full-length TcdB at pH5; it presented a similar structural organization as TcdA. When comparing with the existing crystal structure of

TcdA at physiological pH, only subtle difference is observed around two helices (89). This could be because nanobodies co-crystallized with TcdB limit conformational changes in the delivery domain.

The most hydrophobic region of the TcdB delivery domain was identified by analyzing its primary sequence. Based on this analysis, residues 956 to 1128 are proposed to be the region involved in pore-forming and translocation. The hydrophobic segments in this region can be divided into multiples of 18-25 amino acids, suggesting these hydrophobic patches insert into the membrane as $\boldsymbol{\alpha}$-helices rather than as a $\boldsymbol{\beta}$-barrel. Residues 955 to 990 are proposed to be the minimum sequence required for pore formation. In this region, there are two glutamic acid residues at positions 970 and 976 that are critical for the pore formation as they are proposed as "pH sensor" for the following conformational change. Replacing these residues with alanine prevents pore formation (90). When the pH drops, these residues may become protonated and capable of being inserted into the lipid bilayer (91). It is reasonable to propose that large conformational changes in the four-helix bundle (residues 1026-1135) that are critical for pore formation will occur when the pH drops. When TcdB is in a neutral pH environment, these helices wrap around the delivery domain and keep the protein soluble.

Another analysis compared the sequence of the hydrophobic region of TcdB with the DT toxin; the pattern of hydropathy within the hydrophobic region is conserved within the LCT family. The region between residues 1035 to 1107 of TcdB forms a hydrophobic patch comparable to that of the DT toxin. Based on the results of a site-

directed mutagenesis study, some amino acids in this region were identified as critical for pore formation. By mutating the conserved residues to cysteine or lysine and measuring the ability of the mutant proteins to form pores, it was found that the L1106K substitution has the most drastic effect, totally abolishing the ability to form pores.

Besides determining the critical residues by mutagenic analysis, structural studies of TcdB at acidic pH were also performed over the past few decades. It was proposed that part of the TcdB pore-forming region shares a structure similar to that of diphtheria toxin (DT) during translocation. Four a-helices insert into the membrane and form a so-called "double-dagger" structure (86). When comparing the recent full-length crystal structure of TcdB at pH 5 to the structure of TcdB at physiological pH, the biggest difference was seen in the region of residues 1024 to 1048 (89). Two helices and a loop that connects these two helices that are present at physiological pH are missing in the crystal structure at pH 5, suggesting that this region is flexible. This might correspond to the conformation the toxin assumes when it inserts into the endosome membrane. However, even though much evidence to indicate that TcdB translocation is truly pH dependent, there is no direct evidence to illustrate how this process occurs.

1.3.1.3. TcdA/TcdB autoprocessing

Even though TcdA and TcdB are the largest clostridial toxins, only the N-terminal GTD is released into the cytosol of the host cell. The remainder of the protein is retained on the late endosome and will eventually be digested. The GTD domain is cleaved after a conserved leucine residue (542 in TcdA and 543 in TcdB) (92). This cleavage occurs at neutral pH, suggesting that the translocation of TcdA/TcdB delivers

the GTD together with the APD into the cytosol. Autoprocessing requires inositol hexaphosphate (InP_6), which is abundant in the cytosol of mammalian cells (93). The APD is a cysteine protease that cuts and releases GTD from the other domains of TcdA/TcdB. Mutating the three conserved amino acids D589, H655 and C700 in TcdA or D587, H653 and C698 in TcdB inhibits autoprocessing activity, suggesting that these three residues constitute the autoprocessing pocket. A zinc ion interacts with H758 and D547 and is probably required for the regulation of APD activity. A work demonstrated that APD is in an equilibrium between the active and inactive form prior to binding with InP_6 (94). The coupling with InP_6 significantly shift the equilibrium toward the conformation of active APD. The crystal structure of APD bound to InP_6 demonstrates that InP_6 interacts with a positively charged pocket, which is separated from the active site by a structure called the "$\boldsymbol{\beta}$-flap" (residues 746-765) (95). An allosteric conformational change is observed when the structures of apo-APD and InP6-bound APD are compared (96). The $\boldsymbol{\beta}$-flap rotates about 90^o, causing subsequent movement in the following amino acids (766-802), thus activating the enzymatic site (Figure 1-8). Replacement of H759 with alanine generates an APD that is able to self-process without InP6 binding, suggesting that H759 is critical for the allosteric regulation of the APD domain (96). In a recently published paper, it is showed that the N-terminus of CROPS domain (1792-1834), referred as "hinge", inhibits the autoprocessing activity of TcdB APD domain. This conclusion is supported by the observation that $TcdB_{1-1805}$ has a higher autoprocessing activity than full-length TcdB. A similar result was also obtained with TcdA (97).

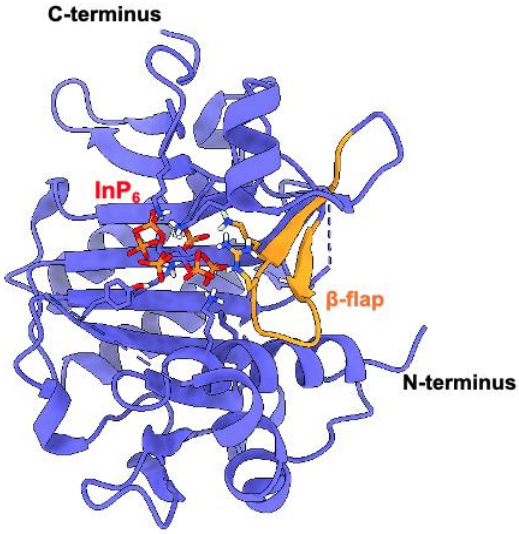


Figure 1-8 The crystal structure of TcdA APD domain.
The N-terminus and C-terminus of the APD is labeled. The structure of β-flap is colored in orange. The InP_6 structure is showed in ball and stick mode.

The autoprocessing activity of TcdB and TcdA is a key factor in the virulence of *C. difficile*. In hypervirulent *C. difficile* strains, TcdB has higher autoprocessing activity and generates more free GTD (98). As a defense mechanism, the host cell can nitrosylate the APD domain of TcdB to block autoprocessing and attenuate the toxicity of TcdB.

1.3.1.4. TcdA/TcdB glucosyltransferase activities

The targets for TcdA and TcdB are small GTPases that are involved in many cellular activities via a wide range of effector molecules. Among these are factors that are involved in the actin assembly of the cytoskeleton. These GTPases have an active

GTP binding form and an inactive GDP binding form. Three proteins are involved in small GTPase regulation. Guanine nucleotide exchange factors (GEF) exchange GDP with GTP on the GTPase to activate the enzyme. GTPase activating proteins (GAPs) are used to stimulate GTP hydrolysis activity to inactivate the signaling activity of the GTPase. Guanine nucleotide dissociation inhibitors (GDIs) are used to extract the GTPases from the membrane to the cell cytosol and maintain the GTPase in its GDP-bound inactive form (Figure 1-9) (99).

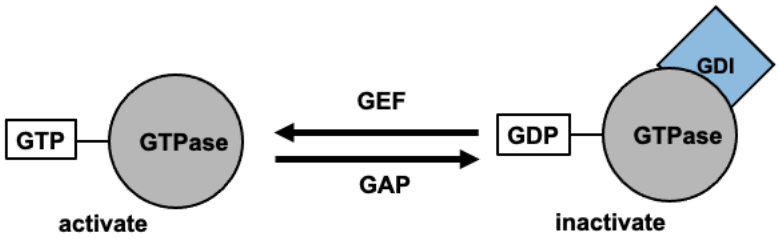


Figure 1-9 Scheme of the function of GTPase regulation factor.

TcdA, TcdB and other LCT toxins all have GTDs that can modify small GTPases irreversibly, thus inactivating their activities and downstream pathways. GDP-bound, membrane-associated forms of GTPases are the preferred substrates for LCT toxins. The free GTD domain released from TcdB is able to monoglucosylate its targets, such as Ras or Rec, on the threonine in the switch I region of the GTPase. The switch I region is critical for binding Mg^{2+} as well as the replacement of hydrolyzed GDP with GTP (100).

Thus, this monoglucosylation disables the regulation of the GTPase and its enzymatic activity. The glucose utilized by TcdB GTD molecule is supplied by uridine diphosphate glucose (UDP-glucose). The crystal structure of GTD shows that the core is a Rossman-fold structure with a D-X-D motif (Asp-X-Asp) that is essential for enzymatic activity (101). This motif is conserved in all LCTs and is important for binding the magnesium cofactor. Mutations targeting this motif significantly decrease enzymatic activity.

Recent research showed that the human chaperone TRiC might be involved in the TcdB glycosylation activity by folding the released GTD domain back into the correct tertiary structure. It is known that TRiC interacts with the GTD domain of TRiC, and knockdown or inhibition of TRiC decreases the toxicity of TcdB in a cell assay, suggesting that TRiC is essential for TcdB toxicity (63). How this happens and where TRiC interacts with TcdB was not elucidated.

1.4. Treatment of *Clostridium difficile* infection

When colonized with *C. difficile*, the patient often shows no symptoms. The onset of CDI is usually associated with dysbiosis of gut microbiome (102). It is important to distinguish asymptomatic CDI from symptomatic CDI for a decision on treatment. To diagnose if the patient is infected by *C. difficile*, a polymerase chain reaction (PCR) is usually performed on a patient's stool sample to see if there is any *C. difficile* gene profile (103). Alternatively, an immunoassay that tests for toxin production may be run (104). Multi-step testing, instead of a single test, is recommended to improve the accuracy of the diagnosis to avoid improper treatment. Once the patient is diagnosed as CDI positive, some routine treatments will be carried out on him.

Around 20-30% patients who suffer from CDI will experience a first recurrence after they have fully recovered (105). Around 50-60% patients will experience secondary recurrences after the first recurrence (106). Different antibiotic treatments result in different recurrence rates. For example, patients treated with oral fidaxomicin experience recurrence at 13%, whereas patients treated with have a recurrence rate of 25% (107).

An antitoxin immune response is the major cause of recurrence. Once *C. difficile* infection starts, the secreted TcdA and TcdB will cause damage and inflammation in the epithelial cells of the large intestine. In order to attenuate the damage, a mucus layer covers the epithelium of the intestine. Also, antimicrobial peptides and S-nitrosylation are used to protect or attenuate the cytotoxicity of TcdA and TcdB by inhibiting toxin cleavage and its entry (108). The recurrence of CDI is related to the speed and extent of the immune response to the toxin. Some people who are infected with *C. difficile* contain high levels of anti-toxin IgG and do not show CDI symptoms, and there is no recurrence of CDI in these patients.

When the patient has a mild to moderate CDI infection, vancomycin, fidaxomicin or metronidazole are used for treatment (109). For recurrent CDI, vancomycin and metronidazole are the antibiotics of choice. To prevent recurrent CDI (rCDI), it is recommended to use vancomycin or fidaxomicin initially, as metronidazole is associated with a higher rate of rCDI (110).

In addition to the traditional antibiotic treatments, there are some novel therapies available to reduce the inflammation as well as the rate of rCDI. New agents are the

antibiotic ridinilazole, live probiotics, bacteriophage treatment and antibody treatment. The human monoclonal antibody bezlotoxumab, under the brand name Zinplava, was recently approved by the FDA. By binding on the CROPS domain of TcdB and sterically blocks the interaction between TcdB and its target receptor , it is able to reduce toxin-induced inflammation in the colon and reduce the rate of rCDI (111).

2. STRUCTURAL ANALYSIS OF *C. DIFFICILE* TCDB INTERACTION WITH RECEPTORS

2.1. Introduction

Clostridium. difficile is a gram-positive, anaerobic and disease causing bacillus that colonizes the human gastrointestinal tract (GI tract). Patients who are infected by *C. difficile* experiences from mild diarrhea to severe pseudomembranous colitis, and such infection is lethal (36). These symptoms are caused by the Large Clostridial Toxins (LCTs) secreted from *C. difficile,* which disrupt host cell skeleton and small GTPase related signaling pathways (45). TcdB is one of the major LCT that are responsible for most CDI symptoms. TcdB is produced and infects host colonic tissues after *C. difficile* germination as the homeostasis of gut microbiota is disrupted by usage of antibiotics. As a multidomain toxin, TcdB can be divided into four domains and they are involved in different steps during TcdB uptake. The N-terminal glucosyltransferase domain (GTD, residues 1-543) is able to irreversibly add a mono-glucose from UDP-glucose onto small GTPase, disrupting the related signaling pathways and cell skeleton assembly. Autoprocessing domain (APD, residues 544-841) is a cysteine protease that can self-process after a conserved cysteine, releasing the GTD into cytosol. Delivery domain (residues 842-1834) extends like a rod shape, with a hydrophobic region, consisting of four alpha-helices, wrapping around the delivery domain. Conformational changes happen in the delivery domain when the endosome lumenal pH decreases, and GTD and APD will be translocated from endosome lumen to cytosol. The combined repetitive oligopeptides (CROPS) domain (residues 1835-2367) points to the opposite side against

the delivery domain, and is proposed as the potential receptor binding domain and is involved in the regulation of autoprocessing activity (72).

Receptor binding of TcdB is the first and the critical step for its entry. Studies have shown that frizzled-2 (FZD2) and chondroitin sulfate proteoglycan 4 (CSPG4) are the receptors for TcdB (69) (80). Knockout both membrane proteins will end up with no toxicity from TcdB to the human cell (81). FZD2 and CSPG4 interaction regions were identified in previous studies, suggesting that FZD2 and CSPG4 interact with TcdB through distinct mechanisms. Crystal structure of FZD2 and TcdB fragments is resolved and provides more detailed information of their interactions (1). However, there is no structural information about the interactions between CSPG4 and TcdB, possibly because building the CSPG4 model is challenging without any homologous structures available. Also, it is interesting to investigate the dynamic differences between the FZD2-bound state TcdB and CSPG4-bound state TcdB. We utilize cryo-EM to resolve high resolution structures of TcdB and CSPG4 fragment complex and TcdB and FZD2 extracellular domain complex. We also combined with other biochemical assays to demonstrate a dynamic state of TcdB when it is interacting with its receptors and provides a thorough of TcdB uptake mechanism.

2.2. Materials and Methods

2.2.1. CSPG4(410-560) preparation

2.2.1.1. pEGFP-N1-CSPG4 reconstruction

The cDNA sequence of CSPG4 (410-560) is synthesized on a pUC17 plasmid. Primers that carry the N-terminal signaling sequence of CSPG4 and restriction enzyme

cutting sites are designed and incorporated in synthesized DNA sequence through two rounds of polymerase chain reaction (PCR). Double enzyme digestion was performed on both CSPG4 DNA and the purified pEGFP-N1 plasmid. This pEGFP-N1 has a C-terminal GFP expressed downstream of the multiple cloning sites (MCS). Then the plasmid and the DNA will be ligated through T4 ligase at the ratio 1:5 under room temperature for 1hr. The ligation mixture will then be transformed into competent DH5α *Escherichia. coli (E.coli)* and four colonies that contain reconstructed plasmid were selected and sequenced for accuracy. The final DH5α that contains the correct CSPG4 sequence on pEGFP-N1 was saved in 10% glycerol at -80 ˚C for storage. The cell will then be inoculated in the Luria-Bertani (LB) medium with 50ug/mL Kanamycin antibiotics overnight for plasmid extraction on the next day.

2.2.1.2. CSPG4(410-560) fragment purification

1mg reconstructed pEGFP-N1-CSPG4 was purified through Qiagen maxi plasmid extraction kit prior to the cells preparation. FreeStyle 293-F cells were seeded into fresh FreeStyle 293 expression media with a final density of 1.0×10^6 cells ml^{-1} and incubated at 37 ˚C, 8% CO_2, 130 RPM. After 24 h, the purified 1 mg pEGFP-N1-CSPG4 plasmids, 2 mg of linear PEI25000 (Polysciences, Inc.) were mixed into 50mL 1 ×Phosphate buffer saline (PBS) and incubated at room temperature for 20 min. Then, the mixture was added into 1L FreeStyle 293-F cells. The transfected cells are incubated at 37 ˚C, 8% CO_2, 130 RPM for 6 days. After 6 days, the cells were pelleted at 3000 RPM for 5 min. One tablet of protease cocktail inhibitor (Roche) was added into supernatant and 2ml streptavidin agarose beads (EMD Millipore) subsequently. The

streptavidin agarose beads have been coupled with biotin-labelled Avi-SUMO-GFP nanobody previously to bind with supernatant. After shaking at 4 ˚C overnight, the beads were pelleted at 500xg for 5 minutes and continuously washed with a pre-chilled washing buffer (20 mM Tris pH 7.5, 150 mM NaCl) for 100mL. Finally, the beads were resuspended in a 2mL washing buffer and SUMO protease was added to cleave the target protein from the beads. To fully elute protein, the beads were then washed again using a washing buffer for another 2mL. The eluted protein was further purified using a superdex 200 increase 10/300 GL column (GE Healthcare) and concentrated using a amicon ultra-4 concentrator (Millipore). The final concentration of purified CSPG4 fragment is at 2mg/mL and was fast frozen using liquid nitrogen before put into -80 ˚C for storage.

2.2.2. CSPG4 (410-560) mutants purification

2.2.2.1. Reconstruction of CSPG4 (410-560) mutants

Primers for CSPG4 mutants were designed through NEBaseChanger (Table 2-1). By following the protocol of Q5 site-directed mutagenesis kit (NEB), the wild type pEGFP-N1-CSPG4 was used as a template to make corresponding CSPG4 mutant plasmids. These plasmids were treated with DpnI to remove the methylated template and then transformed into competent DH5α cells. Three colonies of each mutant were picked after incubating the competent cells on 50ug/mL kanamycin LB agar plates overnight at 37ºC. Plasmids were extracted from these colonies and sequenced for the correct DNA sequences. These plasmids were used for future CSPG4 mutants protein purification.

Table 2-1 Primers for wild type CSPG4 and CSPG4 mutants design.

	Primer name	Primer sequence (5' to 3')
Wild type CSPG4	Forward-1	GCCTTGGCTTTGACCCTGACTATGTTGGCCAGACTTGCATCCGCGGAGCTGC CTGAGCCATGCGTGC
	Reverse-1	CGCGCGCTGCCGCGCGGCACCAGGCCGCCATGTGGGAAGATGATG
	Forward-2	CTAGCTAGCATGCAGTCCGGCCCCCGCCCCCCACTTCCAGCCCCCGGCCTG GCCTTGGCTTTGACCCTGAC
	Reverse-2	CGCGCGCTGCCGCGCGGCACCAGGCCGCCATGTGGGAAGATGATG
Mutant 461E to Q	Forward	GGACCTGATGCAGGCTGAGCTGC
	Reverse	AGCGTGGGCTGCACATGC
Mutant 461E to A	Forward	GGACCTGATGGCTGCTGAGCTGCGC
	Reverse	AGCGTGGGCTGCACATGC
Mutant 458D to N	Forward	GCCCACGCTGAACCTGATGGAGG
	Reverse	TGCACATGCCTCCACTCAAGC
Mutant458D to A	Forward	GCCCACGCTGGCCCTGATGGAGG
	Reverse	TGCACATGCCTCCACTCAAGC
Mutant 451R to A	Forward	GCTTGAGTGGGCCCATGTGCAGCCCACGC
	Reverse	CAGGCTGTGCCCCCCTCG
Deletion 486I and 487P	Forward	GGAGCCCAGGCACGAAAAATG
	Reverse	GTCCAGCTCGAGCTCGCC

2.2.2.2. CSPG4 mutant purification

To express each of the CSPG4 mutants, 0.4mg reconstructed plasmids for each mutants were purified through Qiagen plasmid maxi kit prior to the cells preparation. FreeStyle 293-F cells (Thermo fisher scientific) were seeded into fresh FreeStyle 293 expression media (Thermo fisher scientific) with a final density of 1.0×10^6 cells ml^{-1} and incubated at 37 ˚C, 8% CO2, 130 r.p.m. After 24 hrs, each of the purified 0.4 mg plasmid and 0.8 mg of linear PEI25000 (Polysciences, Inc.) were mixed into 30mL 1 × PBS and incubated at room temperature for 20 min. Then, the mixture was added into 400mL FreeStyle 293-F cells. The transfected cells are incubated at 37 ˚C, 8% CO2, 130 r.p.m for 6 days. After 6 days, the cells were pelleted at 3000 rpm for 5 min. Each collected mutant supernatant was added one tablet of protease cocktail inhibitor (Roche)

and 0.5ml clean streptavidin agarose beads (EMD Millipore) subsequently. The streptavidin agarose beads have been coupled with biotin-labelled Avi-SUMO-GFP nanobody previously to bind with supernatant. After shaking at 4 ˚C overnight, the beads were pelleted at 500g for 5 minutes individually and continuously washed with a pre-chilled washing buffer (20 mM Tris pH 7.5, 150 mM NaCl). Finally, the beads were resuspended in a 1mL washing buffer and SUMO protease was added to cleave the target protein from the beads. To fully elute protein, the beads were then washed again using a washing buffer for another 2mL. The final concentrations of purified CSPG4 mutant fragments were around 0.5mg/mL and was fast frozen using liquid nitrogen before put into -80 ˚C for storage. Among the six mutants transfected in the 293-F cell, only four mutants were expressed. They are: 1). Mutation from 461E to Q; 2). Mutation from 458D to A; 3). Mutation from 451R to A; 4). Deletion of 486I and 487P. These mutants are used for testing the interaction with wild-type TcdB through ELISA assay.

2.2.3. Reconstruction of truncated TcdB (550-1902) and site directed mutagenesis TcdB

Primers are designed based on the sequence of pHis22-TcdB (sequence from VPI 10463 strain) with BamHI cutting sequence in the forward primer and XhoI in the reverse primer. The PCR product was gel verified and purified by QIAquick PCR purification kit (Qiagen) and pasted on pET28a plasmid with a His-sumo tag on the 3' end using T4 ligase (New England Biolabs) at room temperature for 1hr. The mixture was transformed into a competent DH5α cell and the colonies are selected using kanamycin antibiotics at 50ug/mL. The plasmids extracted from four picked colonies

were sequenced and the one that had the correct sequence of TcdB truncation was saved for later use. This newly sequenced reconstructed plasmid was used as template for the following site-directed mutagenesis tcdB mutant variants. Following the instruction of Q5 site-directed mutagenesis kit (New England Biolabs), primers containing the mutated base pair were mixed with a template, together with the Q5 enzyme and buffer. The mixtures were used for PCR and the products were digested using Dpn I enzyme mix for 30 minutes. After transforming the mutated plasmids into DH5α, the plasmids were sequenced and transformed into BL21(DE3) strain for expression.

2.2.4. Purification of TcdB proteins

Plasmid DNA encoding truncated wild type TcdB and its mutants were transformed into BL21(DE3) cells and inoculated to OD_{600}= 0.5 in 1 Liter LB medium. 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added in the flask and incubated overnight at 18 ˚C at 200rpm. The cell pellets were collected and lysed after adding 1 tablet of protease inhibitor cocktail (Roche). The supernatant was collected after centrifuge the lysate at 15,000 rpm for half an hour at 4 ˚C. Pre-washed Ni-NTA affinity column was used for interacting with supernatant for one hour at 4 ˚C before washed extensively with washing buffer (50mM Tris, pH 7.5, 150mM NaCl, 25mM imidazole). The TcdB was eluted using elution buffer containing high concentration of imidazole. The collected TcdB was then concentrated at ~0.5mg/mL and stored at -80 ˚C for future use.

Similar approach was used for purifying wild-type TcdB. Plasmid DNA encoding a 6-His–tagged TcdB was transformed into *Bacillus megaterium* cells, and the

recombinant TcdB was purified via Ni-NTA affinity column essentially as described previously by Yang and colleagues (112). The column was washed with high-salt PBS (20 mM NaH_2PO_4, 20 mM Na_2HPO_4, 300 mM NaCl [pH 7.4]) containing 25 mM imidazole, and the bound protein was eluted using high-salt PBS containing 250 mM imidazole. The eluted protein was then loaded onto a Superdex 200 increase 10/300 GL column (GE Healthcare) and all the fractions are confirmed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

2.2.5. ELISA assays on TcdB

2.2.5.1. Interaction between wild-type TcdB and wild-type CSPG4/mutants

MaxiSorp immuno plates (Nunc) were coated with 4 μg/mL TcdB overnight at 4˚C. The next day, the wells were washed and blocked with PBSTB buffer (PBS containing 0.1% Tween-20 and 2% BSA) before being incubated with serially diluted wild-type CSPG4 (containing an GFP at the C-terminus),or mutated CSPG4. After incubation, wells were washed 4 times with PBST. Bound CSPG4-EC-GFP was detected using rabbit anti-GFP antibody (0.05 μg/mL, Proteintech, Rosemont, IL [catalog #50430-2-AP]) and HRP-conjugated anti-rabbit antibody (0.8 μg/mL, Santa Cruz Biotechnology, Dallas, TX [catalog #SC-2004)]. The color development agent was 3,3’,5,5’-tetramethylbenzidine (TMB).

2.2.5.2. Interaction between mutated TcdB and wild-type CSPG4

Similar approach was used as described above. MaxiSorp immuno plates (Nunc) were coated with 4 μg/mL wild-type CSPG4 overnight at 4˚C. The next day, the wells were washed and blocked with PBSTB buffer (PBS containing 0.1% Tween-20 and 2%

BSA) before being incubated with serially diluted truncated wild-type TcdB or mutated TcdB (containing a His-tag at the C-terminus). After incubation, wells were washed 4 times with PBST. Bound TcdB was detected using rabbit anti-His antibody (0.05 μg/mL) and HRP-conjugated anti-rabbit antibody (0.8 μg/mL, Santa Cruz Biotechnology).The color development agent was 3,3',5,5'-tetramethylbenzidine (TMB).

2.2.5.3. Interaction changes between wild-type TcdB and FZD2/CSPG4 after pH changes

MaxiSorp immuno plates (Nunc) were coated with 4 μg/mL TcdB overnight at 4˚C. The next day, the wells were washed and blocked with PBSTB buffer (PBS containing 0.1% Tween-20 and 2% BSA) before being incubated with serially diluted wild-type CSPG4 (containing an GFP at the C-terminus),or FZD2. After incubation, wells were washed 4 times with PBST. HRP-conjugated goat anti-human antibody (0.025 μg/ mL, Jackson Immuno Research, West Grove, PA [catalog #109-035-088]) was used to detect bound FZD2 and bound CSPG4 was detected using rabbit anti-GFP antibody (0.05 μg/mL, Proteintech, Rosemont, IL [catalog #50430-2-AP]) and HRP-conjugated anti-rabbit antibody (0.8 μg/mL, Santa Cruz Biotechnology, Dallas, TX [catalog #SC-2004)]. The color development agent was 3,3',5,5'-tetramethylbenzidine (TMB). Or the preincubated TcdB with receptors were washed extensively with citric acid buffer at pH5 before binding with correspondent antibodies.

2.2.6. Cryo-EM of TcdB and receptor complexes

2.2.6.1. Cryo-EM TcdB and FZD2

2.2.6.1.1. Sample preparation of TcdB and FZD2

Purified full-length TcdB and FZD2-EC (R&D system) were mixed at 1:1 molar ratio (with the final concentration of the complex at 800 nM) and incubated in PBS buffer at pH 7.4 for 30 minutes at room temperature. 3 μL of the complex was applied to C-flat 1.2/1.3 holey carbon film 300 mesh grids (Electron Microscopy Sciences) at 20˚C with 100% relative humidity and vitrified using a Vitrobot (Mark III, FEI Company, the Netherlands). The grid was then fast frozen into prechilled liquid ethane and then transferred into liquid nitrogen for further storage.

2.2.6.1.2. Cryo-EM data collection

The grid of TcdB and FZD2 complex was imaged under the Titan Krios G3 transmission electron microscope (Thermo Fisher Scientific) operated at 300 kV. The microscope is equipped with a Gatan K2 summit direct detection camera (Gatan, Pleasanton, CA); 5,478 micrographs were collected using electron-counting mode at pixel sizes of 1.06 Å. The beam intensity was adjusted to $7e^{-}/Å^{2}/s$ on the camera. A 30-frame movie stack was collected for each micrograph, with 0.2 seconds per frame, for a total exposure time of 6 seconds.

2.2.6.1.3. Cryo-EM data processing

The collected micrographs were motion corrected using MotionCorr2 (113). The dose-weighted micrographs were visually screened, and 4,920 micrographs with strong power spectra were selected for further processing. Contrast transfer functions (CTF) of

the micrographs were estimated using Gctf (114). Particles were picked using Gautomatch with 2D templates derived from previously published density map (115). These particles were scaled to a pixel size of 4.24 Å by scaling the extracted particles using RELION (116). The automatically picked particles were then screened for high-contrast particles for 4 rounds of the reference-free 2D classification in RELION; 146,129 clean particles were selected and combined for 3D classification, separating particles into 4 classes. All of 81,240 clean particles from 3D classification were used for 3D refinement. The final density map of TcdB and FZD2 is at 5.2 Å. The same particles were also used for 3D refinement of the core region with a solvent mask that masked out the CROPS region, generating a density map of the TcdB–core–FZD2 complex at 5.0 Å resolution. The overall resolution was assessed using the gold-standard criterion of Fourier Shell Correlation, with a cutoff at 0.143, between 2 half maps from 2 independent half-sets of data. Local resolutions were estimated using Resmap (117).

Multibody analysis was performed using RELION-3.0 on the dataset of TcdB and FZD2 complex. The density map was divided into three bodies. Each of the body was done principal component analysis. A 3D classification was done on the clean particles, separating particles into four classes by applying a solvent mask only around the tip of the delivery domain when setting the angular search to zero. The final result was that all the particles converged into one class, suggesting that there was only one detectable status of the delivery domain.

2.2.6.1.4. Model building

The 5.0Å density map was used for TcdB-FZD2 model building. We chose the TcdB model we built in our previous published paper (115) as the initial model for TcdB and the crystal structure of frizzled family protein 7 (FZD7-CRD) as the initial model for FZD2-CRD. The homology models for the core region (residues 1–1799) and the CROPS domain of TcdB were rebuilt using swiss-model by posting crystal structure of TcdA (PDB:4R04) and crystal structure of TcdA CROPS domain (residues 1,834–2,101,PDB: 4NP4). Similar approach was used to build the initial model of the FZD2-CRD. These initial models were combined and docked into the EM density map we got UCSF Chimera (118) and refined into the cryo-EM density map using Molecular Dynamics Flexible Fitting (119) to generate the complex structure of TcdB and FZD2-CRD. This model was further refined using Phenix (120) to refine and validate the model.

2.2.6.2. Cryo-EM TcdB and CSPG4

2.2.6.2.1. Cryo-EM sample preparation

Purified TcdB and CSPG4 were mixed at 1:1 molar ratio (with the final concentration of the complex at 800 nM) and incubated in a PBS buffer at pH 7.4 for 30 minutes at room temperature. 3 μL of the complex was applied to C-flat 1.2/1.3 holey carbon film 300 mesh grids (Electron Microscopy Sciences) at 20˚C with 100% relative humidity and vitrified using a Vitrobot (Mark III, FEI Company, the Netherlands). The grid was then fast frozen into liquid ethane and then transferred into liquid nitrogen for further storage. To avoid the preferred orientation in the sample, I added 0.01% DDM

into TcdB and CSPG4 mixture at 2mg/mL for complex concentration. The 3uL of the mixture was applied to C-flat 1.2/1.3 holey carbon film 300 mesh grids as mentioned above.

2.2.6.2.2. Cryo-EM data collection

The grid of TcdB and CSPG4 complex was imaged under the Titan Krios G3 transmission electron microscope (Thermo Fisher Scientific) operated at 300 kV. The microscope is equipped with a Gatan K2 summit direct detection camera (Gatan, Pleasanton, CA); 17,206 micrographs were collected using electron-counting mode at pixel sizes of 1.06 Å. The beam intensity was adjusted to $7e^{-}/Å^{2}/s$ on the camera. A 30-frame movie stack was collected for each micrograph, with 0.2 seconds per frame, for a total exposure time of 6 seconds.

2.2.6.2.3. Cryo-EM data processing

The collected micrographs were motion corrected using MotionCorr2 (113).The dose-weighted micrographs were visually screened, and 16,135 micrographs with strong power spectra were selected for further processing. Contrast transfer functions of the micrographs were estimated using Gctf. Particles were picked using Gautomatch with 2D templates derived from previously published density map (115). These particles were scaled to a pixel size of 4.24 Å by scaling the extracted particles using RELION (116). The automatically picked particles were then screened for high-contrast particles for 4 rounds of the reference-free 2D classification in RELION; 383,721 clean particles were selected and combined for 3D classification, separating particles into 4 classes. All of 206,061 clean particles from 3D classification were used for 3D refinement. The final

density map of TcdB and CSPG4 is at 4.6 Å. The same particles were also used for 3D refinement of the core region with a solvent mask that masked out the CROPS region and tip of the delivery domain, generating a density map of the TcdB–core–CSPG4 complex at 4.2 Å resolution (Figure 2-1). The overall resolution was assessed using the gold-standard criterion of Fourier Shell Correlation (FSC), with a cutoff at 0.143, between 2 half maps from 2 independent half-sets of data. I also used cryoSPARC to do the 3D refinement of the complex using the same batch of data, and ended up with a density map with consensus resolution at 3.7Å. Local resolutions were estimated using Resmap.

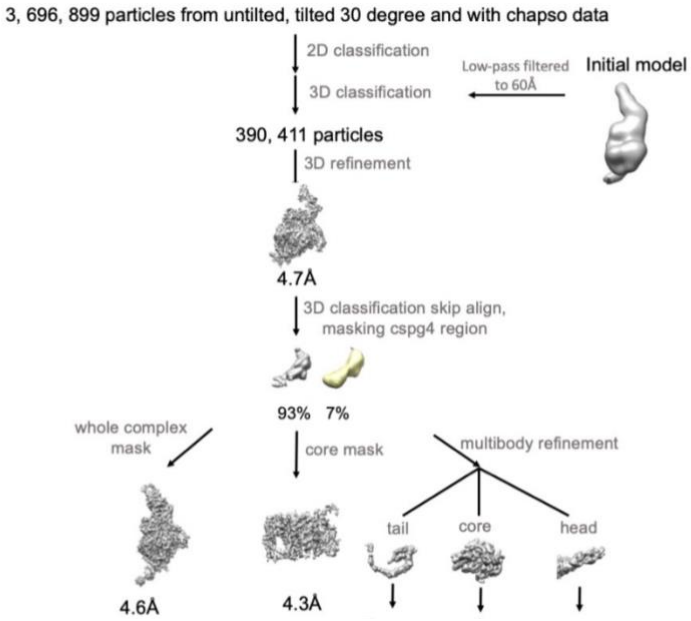


Figure 2-1 The data processing flow chart of TcdB-CSPG4 complex. 3,696,899 clean particles were used to make the consensus map at 4.7Å. By separating the particles around the CSPG4 region, particles from better resolved class were used to generate a better resolved consensus map. When masking around the "core" region, a 4.3 Å map was generated and used for following model building.

Multibody analysis was performed using RELION-3.0 on the dataset of TcdB and CSPG4 complex. The density map was divided into three bodies. Each of the body was done principal component analysis. A 3D classification was done on the clean particles, separating particles into four classes by applying a solvent mask only around the tip of the delivery domain when put “skip align” as one of the settings. The final result showed three states of the delivery domain of the complex. The particles corresponding to each of the class were used for 3D refinement and three density maps were generated, representing the three states of the complex.

2.2.6.2.4. Model building

The 3.7 Å density map was used for TcdB-CSPG4 regional model building. We chose the TcdB model we built in our previous published paper (115) as the initial model for TcdB. This model was modified based on the density map of the TcdB part and deleted the residues that are not included in the density map. The initial model for CSPG4 was generated through Robetta webserver and manually modified using coot (121). Then the initial model of CSPG4 was used in RosettaCM (122) to create 1000 refined models based on two criteria: 1) The overall energy was the lowest of the output models. Needed to fit the Monte Carlo method; 2). The output models need to have the best fit with the cryo-EM density map. The modified TcdB and CSPG4 were roughly fitted in the density map using UCSF Chimera and generated as a single initial model. Then the model was refined into the cryo-EM density map using MDFF (119) to

generate the complex structure of TcdB and CSPG4. This model was further refined using Phenix (120) to refine and validate the model.

2.2.7. Circular dichroism (CD) spectroscopy

The secondary structures of the truncated wild type TcdB and the mutants were analyzed on circular dichroism spectroscopy (Chirascan). All of them were diluted in 10 mM potassium sulfate and 25mM PBS, pH 7.4, to a concentration of 0.1 mg ml^{-1}. After the machine was purged with nitrogen at 20 °C for 40 min, samples were loaded into a 2 mm path length cuvette and the circular dichroism (CD) of the sample measured in the far ultraviolet region. The sample temperature was equilibrated in the spectrometer to 25 °C before the initiation of measurements and was maintained at this temperature throughout. The CD signal at each wavelength was averaged for 30 s, using 1 nm wavelength steps. Scan of the sample buffer was used as blank.

2.3. Results

2.3.1. FZD2 interacts with TcdB at the delivery domain

2.3.1.1. FZD2 grips on the middle of TcdB delivery domain

By using reference-free 2D classification implanted in RELION-3.0 software, we ended up having crispy and clear projection views of TcdB and FZD2 complex. The particles belong to these clear illustrated 2D classes (Figure 2-2A) were selected and processed for 3D classification and refinement using an apo-state TcdB density map at 40 Å resolution to avoid model bias. The finial 3D refinement of TcdB and FZD2 complex is at 5.0Å (Figure 2-2B) with a clear indication of the density of each secondary

structures and some groove of helixes. After fitted and refined model based on the density map, I am able to analyze the interaction between FZD2 and TcdB.

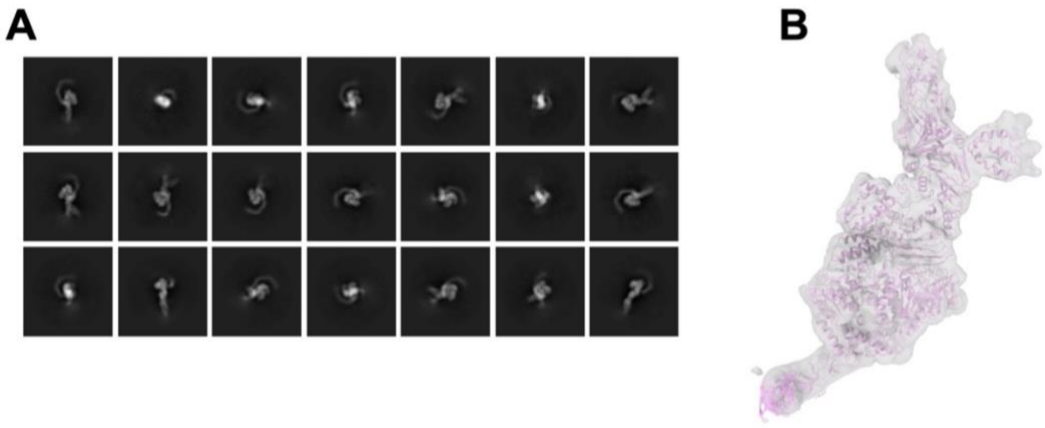


Figure 2-2 2D classification result and 3D refinement of TcdB and FZD2 complex. Panel A shows the 2D classification result of TcdB-FZD2 complex with different views showed. The density map of TcdB-FZD2 is showed in transparent in panel B with the reconstructed model fitted in.

Cryo-EM density shows the full-length of TcdB, with less elucidated delivery domain tip and the C-terminal of CROPS domain, probably because of the flexibility of these two regions. FZD2-cysteine rich domain (CRD) grips on the protrusion of delivery domain. FZD2-CRD is mostly alpha-helices, with its extracellular face touching the TcdB delivery domain. The binding site of FZD2 is novel and different from previously published receptor binding site of TcdB (69). As FZD2 has short cytoplasm domain,

binding of the FZD2 helps TcdB to pull it close to the endosome membrane. This might help with the following pore forming and translocation step.

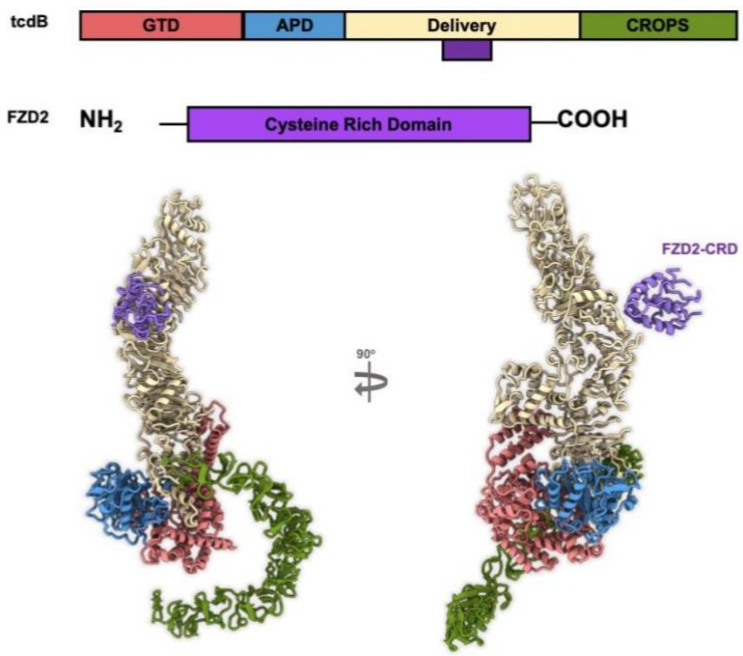


Figure 2-3 model of TcdB and FZD2 complex.
The domain organization and color scheme of TcdB and FZD2 are showed on the top panel, with the correspondent color pattern showed in the model of the complex. The position of FZD2-CRD is highlighted with its name labeled on the side.

2.3.1.2. FZD2 interacts with TcdB through hydrophobic interactions

The overall structure resolution is at 5Å, with a good separation of alpha-helices, some indication of the helix pitch in better-resolved region. Using the homolog models

from TcdA as well as FZD7-CRD crystal structures, I was able to rebuild the model navigated by my density map (Figure 2-3), even though the exact positions for the side chains are not clear, but the relative positions of secondary structures are able to be identified.

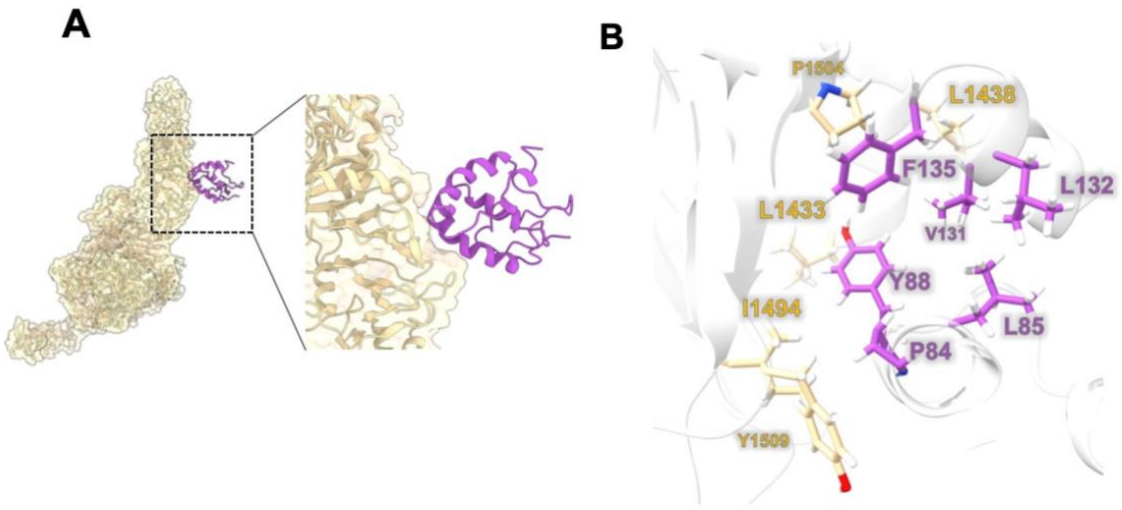


Figure 2-4 Interactions between TcdB and FZD2.
The side chains from TcdB are colored in yellow and the side chains from FZD2 are colored in purple. The residues that are involved in the interactions are highlighted by their amino acid name and their sequence in the protein.

The distance that general strong interactions exist is supposed to be within 5 Å. These interactions includes salt-bridge interaction as well as the hydrophobic interaction. I choose the C-α distance between the two adjacent residues to be less than 5 Å when considering the binding force between TcdB and FZD2. Several residue pairs were found to be within this range. For example, the Proline 1504 and Leucine 1438 in TcdB

were close to the Phenylalanine 135 in the FZD2. The Leucine 1433 and Isoleucine 1494 were close to the Tyrosine 88 in the FZD2. By analyzing the interaction types among these amino acids, I proposed that TcdB interacts with FZD2 through hydrophobic interactions (Figure 2-4). This propose was also confirmed by a recently published crystal structure of partial TcdB in complex with the FZD2-CRD, showing similar interaction pattern with our cryo-EM result. Mutagenesis on the hydrophobic amino acids into the hydrophilic amino acids, the interactions of these two proteins were significantly decreased (89).

2.3.2. CSPG4 interacts with TcdB at the N-terminal of CROPS domain and partial APD domain

2.3.2.1. CSPG4 binds to the ditch formed by the N-terminal of CROPS domain and APD domain

Similar approach was used to get cryo-EM density map of TcdB and CSPG4 complex. Since TcdB and CSPG4 complex has preferred orientation when analyzing data, I tried to tilt-stage to 30^{o} when collecting data and adding detergent n-Dodecyl-B-D-maltoside (DDM) to compensate the preferred orientation effect. As TcdB is very flexible, especially on the CROPS domain and the tip of delivery domain, the reconstruction of the holo-TcdB and CSPG4 is at 4.7 Å (Figure 2-5), with a good separation of secondary structures. In some better resolved region, bulky side chains can be observed. When applying a solvent mask around the more rigid region of the complex, I ended up getting a density map at 3.7Å, with a much more clear side chain

resolved in the TcdB GTD domain and the APD domain, as well as the CSPG4 region. This significantly helped with the following model building.

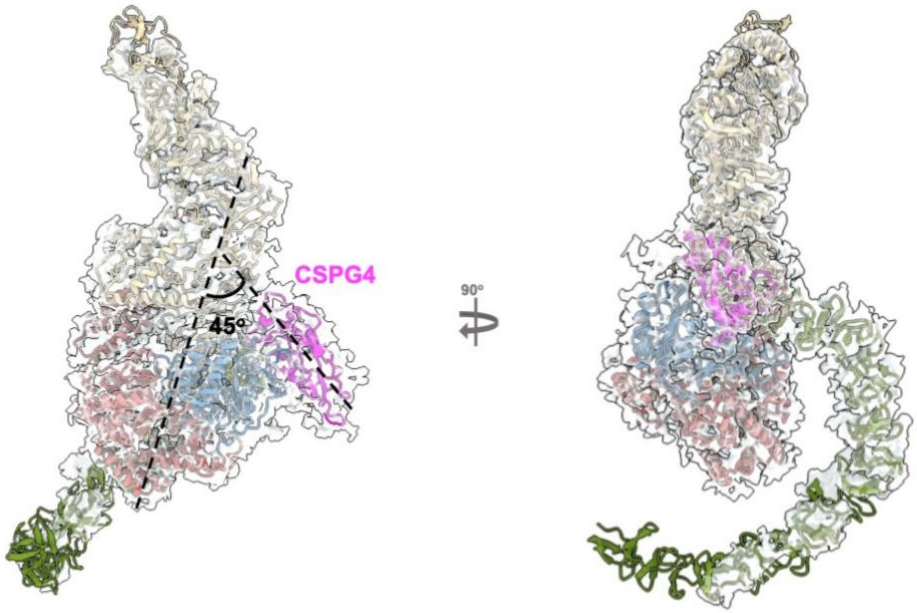


Figure 2-5 Density maps for TcdB and CSPG4 complex.
From the density map of the full-length TcdB and CSPG4, CSPG4 grips on the cleavage formed by the N-terminus of CROPS domain and the APD domain. As an elongated shape, CSPG4 adopts a narrow beta-barrel structure, and forming an angle around 45º with TcdB long-axis.

2.3.2.2. CROPS domain loop is critical for the CSPG4 binding

I tried to get a reasonable initial model before refining it into density map. As there was no homolog model for CSPG4, it is hard to get a decent initial model from I-tasser (123) or swiss-model. I eventually got initial model from Robetta web server that

can roughly fit in our density map (124). By using coot to manually fit in the density map and Rosetta to rebuild the model (Figure 2-6A, Figure 2-6B), I got the structure of CSPG4. There are two verifications of this structure. 1). The orientation of CSPG4 relative to TcdB. Based on the CSPG4 expression plasmid organization, there was a GFP co-expressed on the C-terminus of CSPG4. By applying a large solvent mask around the CSPG4 region, I did 3D classifications to see if I was able to find an extra density that belongs to GFP. The result demonstrated that GFP density was close to the CROPS domain end, indicating that the C-terminus of CSPG4 was toward CROPS domain (Figure 2-6D). 2). A glycosylation site was able to be seen in the density map of CSPG4. As mentioned in many reviews, the CSPG4 is highly glycosylated on its protein surface (125). By predicting the glycosylation site on the 410-560 fragment, one N-linked glycosylation site was found on the asparagine 428 on the position asparagine 428 residue (Figure 2-6C), which matches with our prediction result. From these two evidence, I was able to verify our model correctness.

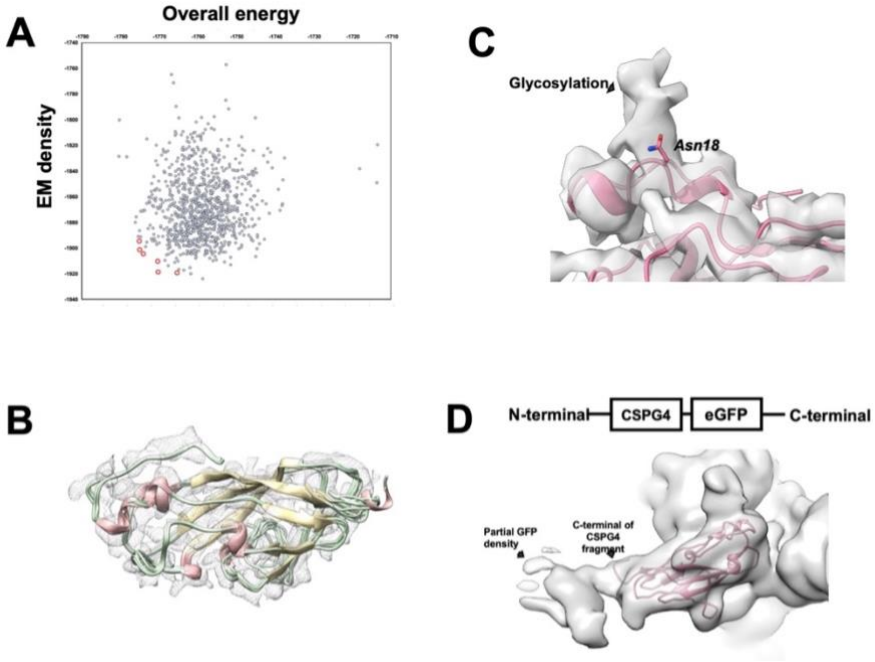


Figure 2-6 Verification of CSPG4 model.
Rosetta build 1000 models of CSPG4 based on the EM density map and the overall energy of the model. Six models that have the best fitting score and the lowest energy were selected for the following refinement. The extra density of glycosylation and the partial GFP density are showed in panel C. Asn18 indicates the glycosylation is on the 18th residue when 410 residue is the N-terminus.

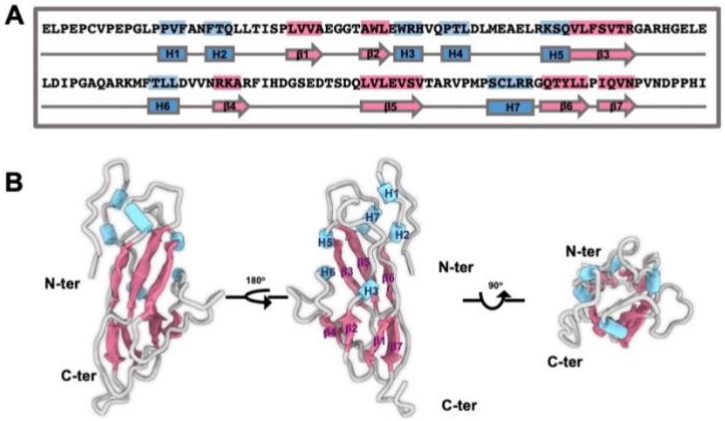


Figure 2-7 The model of CSPG4.
The sequence of CSPG4 is showed in panel A. The beta-sheets are colored in red and the helixes are colored in blue. In panel B, three views of the CSPG4 model are showed, with N-terminus and C-terminus indicated.

The sequence and the model of CSPG4 fragment showed in ribbon style are listed in the figure below (Figure 2-7). α-helices are colored in blue and β-sheets are colored in red. It is showed that the core of CSPG4 fragment is consisted of β-sheets folding next to each other, forming a narrow β-barrel. The β-sheets are connected by short helices and flexible loops.

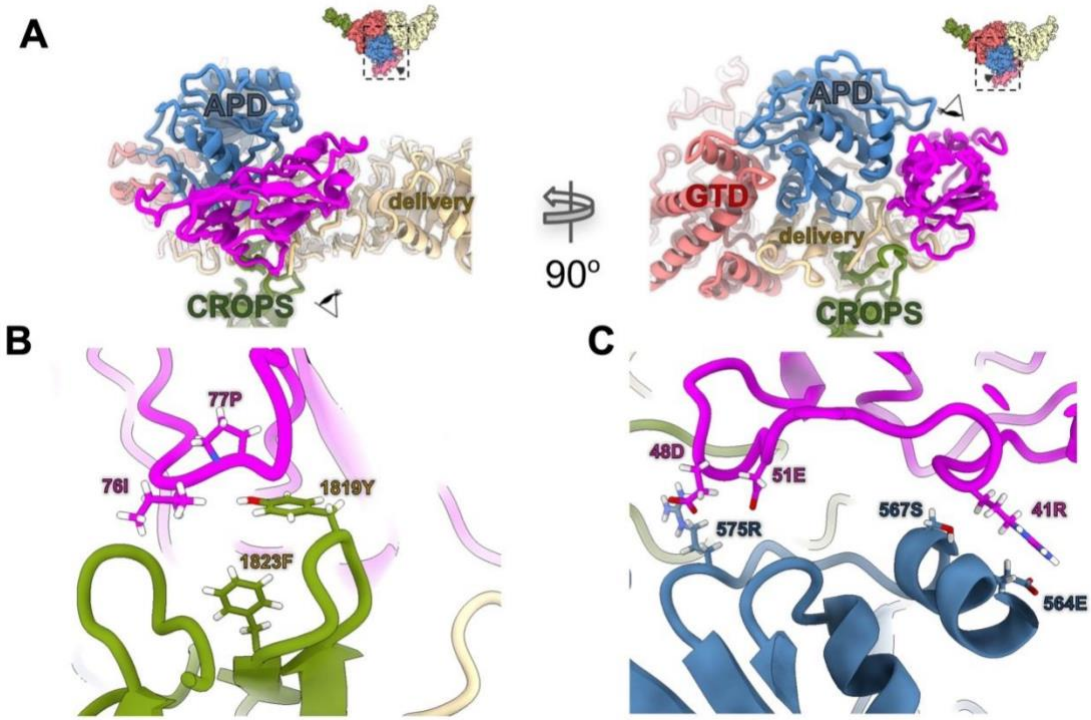


Figure 2-8 Interaction between CSPG4 and TcdB.
The APD is colored in blue. The CROPS is colored in yellow and the CSPG4 is colored in pink. The residues that are involved in the interaction are labeled and with side chains showed.

As the density of the core-region of the TcdB and CSPG4 complex (including GTD, APD, part of CROPS and the CSPG4) was resolved at 3.7 Å, the side chains of residues were much better resolved. Thus I was able to identify the interactions of TcdB and CSPG4 more confidently. CSPG4 touched TcdB at both APD and the CROPS domain. 1). The interaction between APD and CSPG4 was charge-charge interaction. Two pairs of interactions are found in the structure. Residue 575 arginine from TcdB was proposed to interact with 461 glutamic acid or 458 aspartic acid from CSPG4. 451 arginine from CSPG4 was proposed to interact with 567 serine or 564 glutamic acid from TcdB. Their side chains stick together and tends to form salt-bridge in between (Figure 2-8A). 2). The interaction between CROPS domain and CSPG4 is hydrophobic interaction. It is noteworthy that a loop (sequence between β-sheet 3 and α-helix 6) is extending from CSPG4, getting approximate to the CROPS domain. Two hydrophobic residues that contain benzol rings (1819Y and 1823F) were close to this loop, with 486I and 487P as potential interaction sites. These residues form hydrophobic patch around the beginning of CROPS domain (Figure 2-8B).

To verify my propose on TcdB and CSPG4 interaction, I did mutagenesis on the speculated residues on both TcdB and CSPG4. The interactions were measured using ELISA assay after the proteins were purified. The secondary structures of the truncated TcdB and the mutated TcdB were verified using circular dichroism (CD) spectrometry. By mutating the TcdB on the residues described above, we can see that mutating all of the residues on TcdB significantly decreased the interaction. Especially when mutating residue 1822 lysine to alanine. When mutating CSPG4 on the residues described above,

it is showed that the interaction between 451 arginine from CSPG4 with the 564 glutamic acid from TcdB. When mutating 451 arginine into alanine, which didn't have any positive charge or polar on the side chain, the interaction was significantly decrease (Figure 2-9). For the CSPG4 mutant that deleted 486 isoleucine and 487 proline, its interaction with TcdB was also very weak (Figure 2-10). Suggesting the hydrophobic interaction proposed above is true.

Previous result suggested that CSPG4 interacts with TcdB at the N-terminus of CROPS domain, while no interaction on the APD and CSPG4 was reported. Here we combined our high-resolution structure as well as ELISA assay result, suggesting that there was charge-charge interaction present in on the other side of interaction surface on TcdB with CSPG4.

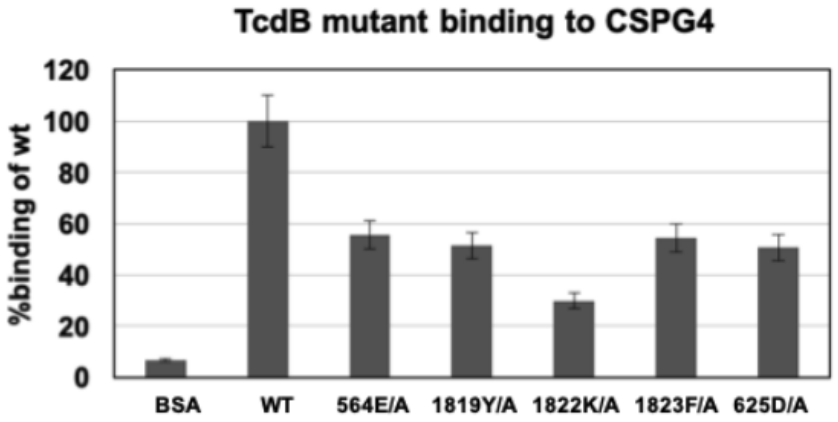


Figure 2-9 The interaction between mutated TcdB and CSPG4 through ELISA. The signal from WT is used as reference and defined as 100%. All the other mutants signals are used to compare with signal from WT TcdB. The assay is repeated three times.

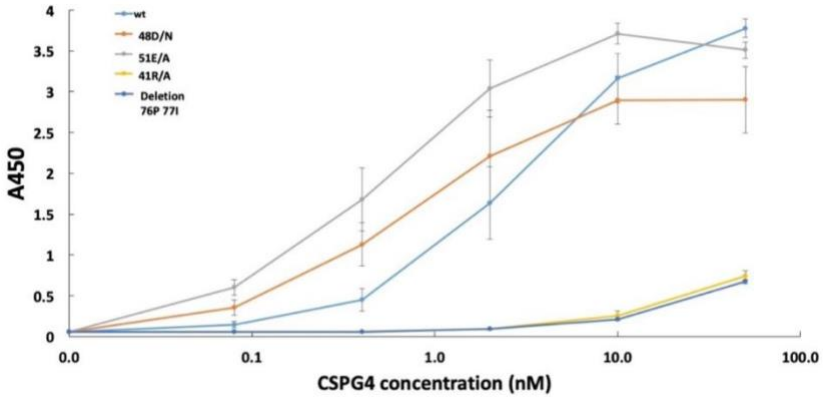


Figure 2-10 ELISA assay on interaction between wild-type TcdB and CSPG4 mutants.
The ELISA assay is done when binding with different concentration of CSPG4. Signals from different CSPG4 mutants are colored based on the legend.

2.3.3. Different receptors have different impact on the stability of TcdB

2.3.3.1. FZD2 stabilizes the structure of TcdB delivery domain when comparing with CSPG4 binding states.

Whether binding with different receptors influences the stability of TcdB is interesting to investigate. In order to visualize the influence on TcdB structure, especially in CROPS domain, we did multibody refinement on both EM density maps we described previously and 3D classification and subsequent refinement was conducted based on the multibody refinement result.

The TcdB and receptor complex density map was divided into “three bodies”. The first body included the density of delivery domain; The second body included the globular density consisted of APD and GTD; The third body included the CROPS domain and the connection between CROPS domain and APD domain. Movements were found on the Delivery domain and the CROPS domain on both density maps. Thus a 3D classification was done when applying a solvent mask around the delivery domain and disable the alignment when running the program (Figure 2-11). This allows RELION to separate particles based on the difference at delivery domain.

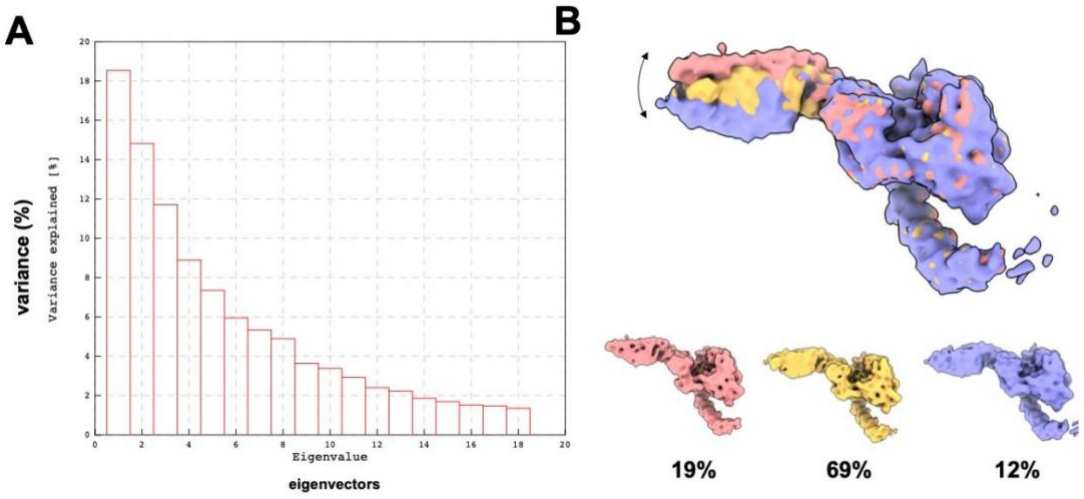


Figure 2-11 The 3D classification of TcdB and CSPG4 complex.
The three classified densities are colored in red, yellow and purple, with their particle percentages labeled at the bottom.

All the particles are separated into four classes and the classification result were compared in Chimera. In theory, if receptor binding state TcdB has distinct variant conformations, I am able to classify them out by doing skip-align 3D classification. For the result from TcdB and FZD2 complex, all the particles converged into one class, indicating that when binding with FZD2, TcdB conformation is fairly conserved while there are three classes were resolved in the TcdB and CSPG4 complex, indicating that CSPG4 is not able to stabilize the TcdB structure. By comparing these three classes, the biggest difference is at the delivery domain tip, where it tends to bend up or down for around 15^{o} (Figure 2-11).

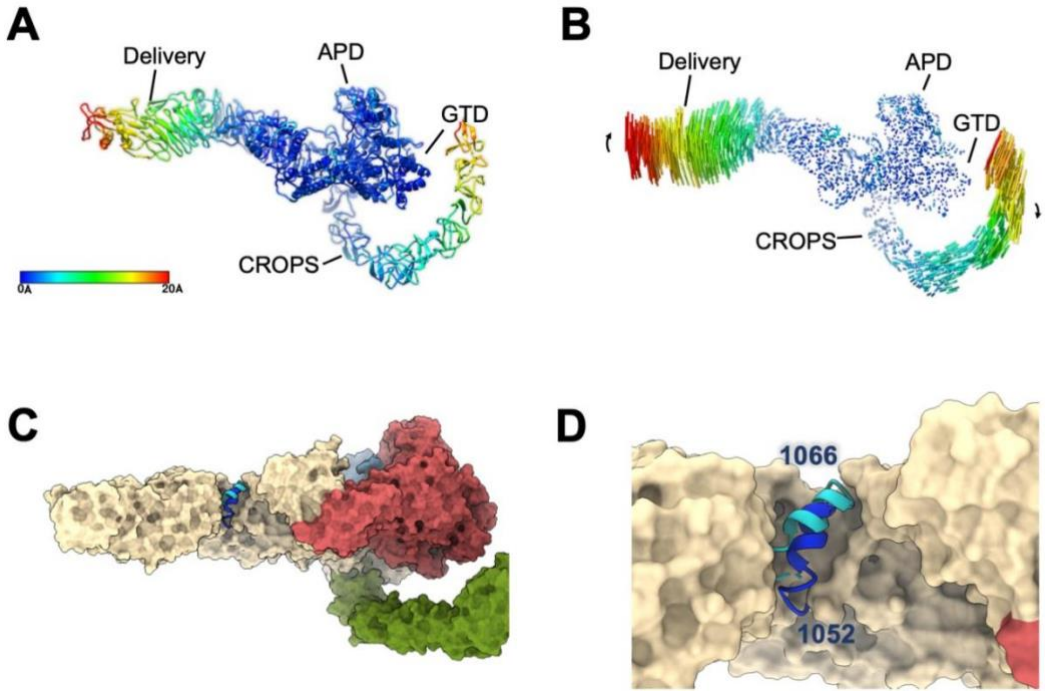


Figure 2-12 The movement of helix 1052-1066.
The movement of the TcdB is colored in panel A and panel B. In panel C and panel D, the position of the helix 1052-1066 is illustrated in deep blue or light blue.

When zooming around the delivery domain, not only the movement of the delivery domain tip is observed (Figure 2-12A, Figure 2-12B), a missing density at two helixes around the hydrophobic region is observed too. When TcdB is in the purple conformation, the delivery domain tip is in the most bent form, while the density of helix 1052-1066 is missing (Figure 2-12C, Figure 2-12D). When TcdB is in the red and yellow conformations, the delivery domain tip is toward up, the density of helix 1052-1066 has a significant movement together with the delivery domain movement. This suggest that the delivery domain of TcdB is flexible. Previous published apo-TcdB cryo-EM density reveals a bent delivery domain tip comparing with the crystal structure of TcdA. When TcdB is interacting with CSPG4, the binding of the receptor doesn't restrict the movement of delivery domain. Making it possible to flipping up and down. The pore-forming related helixes, such as the helix 1052-1066 is unstable too. This flexibility benefits the following conformational change as the delivery domain is at "ready" state by moving around.

2.3.3.2. The two receptors behave differently under acidic conditions when binding with TcdB.

To help understand the TcdB behavior with its receptors, we analyzed the binding affinity of TcdB and these two receptors using ELISA assay. When incubating the TcdB and receptor complex in the physiological pH, the binding strength is defined as 100%. While comparing with the signal from complex under acidic pH, it is

noteworthy that the binding signal from FZD2 stays the same while the signal from CSPG4 decreases significantly (Figure 2-13). This indicates that FZD2 and CSPG4 adopt different interaction pattern. Even though FZD2 and CSPG4 both interact with TcdB through hydrophobic interactions, the interaction surface was different. The hydrophobic patch on the FZD2 interface is fully buried in the TcdB binding surface, thus it is hard for proton or water molecule to get into the interaction interface. While CSPG4 interacts with a loop from N-terminus CROPS domain through hydrophobic interaction, which can be easily accessed by the proton or water when pH drops. Besides, a potential charge-charge interaction was demonstrated at 451R from CSPG4 and 564E from TcdB. When pH decrease, the glutamic acid tends to be more protonated, which will influence with the stability of the salt-bridge interaction.

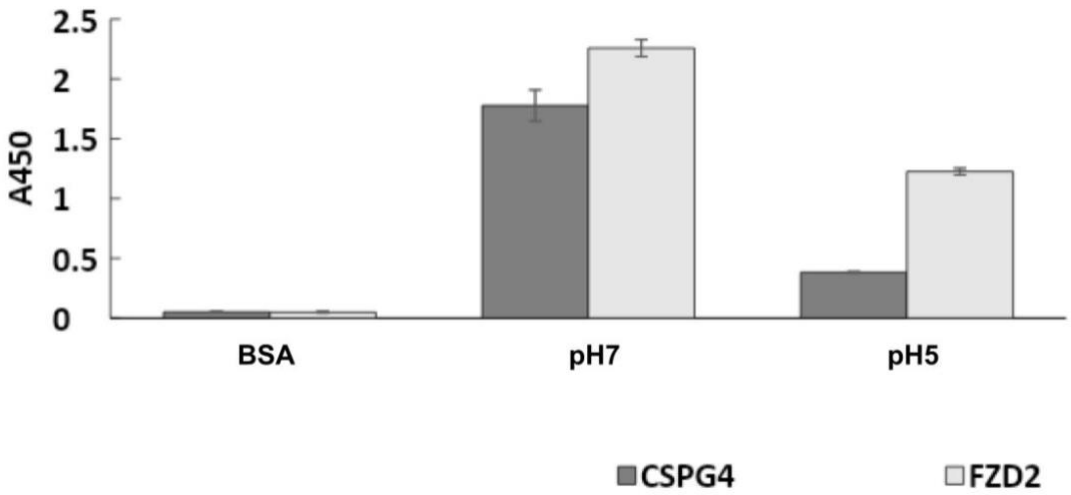


Figure 2-13 Binding assay of TcdB with CSPG4 or FZD2 under different pH.

2.4. Discussion

Like all AB toxins, TcdB uptake into the epithelial cells in the intestine is the first step and the most critical step for its cytotoxicity. TcdB gets into its target cell through Catherin-mediated endocytic pathway after binding on its receptors. Unlike TcdA, which interacts with the poly-saccharide on the cell surface, either on the lipid or on the membrane protein, TcdB has its novel cell entry receptors. PVRL3, FZD2 and CSPG4 are proposed to be the three receptors that mediate TcdB entry. PVRL3 and FZD2 are both expressed in the intestine epithelial cell while CSPG4 is expressed in the submucosa layer, which provides access for TcdB when the surface epithelium is disrupted.

Researches were performed on the interacting sites of the TcdB and its receptors. By doing mutagenesis and truncations on TcdB, it is known that FZD2 and CSPG4 interact with TcdB in different positions. But the detailed interaction surfaces are not analyzed, and no evidence shows the influence on TcdB when binding with the receptors. My studies reveal the two binding states of TcdB with FZD2 as well as CSPG4, providing the details binding surface of these two receptors. FZD2 interacts with TcdB in the middle of delivery domain through a hydrophobic interaction. The CSPG4 interacts with TcdB through more complicated mechanism. When it grips on the ditch between N-terminus CROPS domain and the autoprocessing domain, it interacts with these two domains through different mechanism. By analyzing cryo-EM structure and doing mutagenesis on both TcdB and CSPG4, we understand that CSPG4 interacts

with TcdB autoprocessing domain through charge-charge interaction while the CSPG4 interacts with TcdB CROPS domain through hydrophobic interaction.

When the TcdB goes through the endocytic pathway, it will go through the pH changes from physiological pH to acidic pH, which triggers conformational changes. By analyzing the interactions of TcdB with the two receptors under both physiological pH and acidic pH, it is showed that FZD2 interaction is not influenced by the pH change while CSPG4 will release bound TcdB when the pH decreases. When comparing the extracellular domains of FZD2 and CSPG4, it is easy to tell that CSPG4 has a fairly large and flexible organization. While interacting with CSPG4 on its N-terminal domain, it is hard for TcdB to get close to the endosome membrane and create holes on the endosome membrane. But FZD2 has a globular and small extracellular domain, interacting with FZD2 will help TcdB to be hold approximate to the endosome membrane. Our ELISA assay result provides a logical evidence for the following TcdB translocation step.

By doing 3D classification without performing any alignment on both TcdB binding state data, we found that interacting with FZD2 is actually stabilizing the delivery domain movement while TcdB delivery domain flipping up and down when interacting with CSPG4. It is also noteworthy that among the three states of TcdB when binding with CSPG4, one of the state actually has a missing density around the helix 1052-1066, which is involved in the hydrophobic region of the delivery domain, related to the pore-forming in the following step. Thus I propose that under physiological pH, TcdB itself has a very flexible structural organization, with a flexible CROPS domain as

well as the delivery domain. Only when pH acidified the TcdB will try to stabilize and find a right position to perform the insertion and translocation.

The results described here provides zoomed-in understanding of TcdB receptor binding state and analyzed interacting with receptor might have some effects on the TcdB conformations.

3. STRUCTURAL ANALYSIS OF *C. DIFFICILE* TCDB CONFORMATIONAL CHANGE UPON PH ACIDIFICATION

3.1. Introduction

Clostridium difficile is an anaerobic, gram-positive pathogenic bacillus that causes severe gastrointestinal tract infections, which can be fatal (36). TcdB is one of the major virulence factors that are responsible for *C. difficile* pathogenesis (101). The TcdB N-terminal domain is a glucosyltransferase that can irreversibly add a single glucose from UDP-glucose to membrane-attached small GTPases such as Rho, Rac and Cdc42 in the mammalian target cell (54). This modification will disable the function of these GTPases, blocking the dynamics of the cytoskeleton and the signaling pathways that require these GTPases. TcdB consists of four domains: the glucosyltransferase domain as mentioned above, the autoprocessing domain (APD), the delivery domain and the CROPS domain (101). These four domains have distinct functions for TcdB entry and function in host cells. The delivery domain and the N-terminus of the CROPS domain are the binding moiety that interact with their destinated receptors. After interacting with receptor on the host epithelial cell, TcdB is taken into the cell through endocytosis and into the early endosome. When the pH decreases in the endosome lumen, there will be conformational changes in TcdB, which leads to pore-formation on the endosome membrane as well as translocation of the GTD and APD (72). Only when the GTD and APD are translocated TcdB is able to glucosylate its target. Creating holes and delivering the N-terminal domain is essential for the final modification step of TcdB. TcdB shares similar features with other well-studied AB toxins, such as

diphtheria toxin (DT) and anthrax toxin. Anthrax toxin is secreted from the pathogenic bacterium *Bacillus anthracis*. It has the enzymatic component needed to be translocated into cytoplasm from within the late endosome. It has been reported that the translocation process involves large conformational changes to transform the toxin complex from a pre-pore state into a translocation state. There is evidence indicating that the membrane leakage occurs during translocation and the translocation state of anthrax toxin is well-resolved. At translocation state, anthrax toxin forms a transmembrane segment in a β-barrel (83). However, for all of the toxins that are similar to TcdB, named the Large Clostridial Toxin family, the transmembrane region is formed by α-helices (126). Diphtheria toxin, which utilizes α-helices as its transmembrane region, has the most similar hydrophobicity pattern with TcdB. It is consistent with the three hydrophobic patches. The most hydrophobic helices (named TH8 and TH9) are proposed to insert in the endosome membrane when pH changes(ref). When comparing the primary sequences of the pore-forming domain among different LCT toxins against diphtheria toxin, it is noteworthy that three hydrophobic patches are observed. For TcdB, the three patches are: 1018-1056, 1064-1089, 1091-1112 (86). The loop between helix 960-969 and helix 981-994 is related with the pore-forming because it contains two glutamic acid in position 970 and 976 (127). Mutating these residues result in loss of pore-forming ability of TcdB. The author proposed that these two residues serve as "pH sensor" when the endosome lumen is acidified. The detailed mechanism of this LCT family toxin pore-formation as well as translocation is poorly understood. Since LCT toxins need to exist both soluble at neutral pH while containing hydrophobicity at acidic pH, the membrane-

insertion region should not be too hydrophobic as the canonical membrane protein, which results in instability in the structure when insertion happens. Another line of evidence is that TcdB is shown to form a transient channel after insertion since TcdB insertion cannot generate constant current across the lipid bilayer when interacting with a Rb^+ loaded liposome (128). Thus, the studies about the mechanism of LCT family toxin pore-formation and translocation are only about identify the region that involved in this procedure.

Other than the three hydrophobic patches are involved in the pore-formation, the crystal structure of TcdB in complex with three nanobodies at pH 5 demonstrated that residues 1032-1047 are not visible, suggesting flexibility in this region (1). In combination with the mutations and truncations of TcdB while performing the leakage assay and the toxicity assay results, four helixes in the TcdB hydrophobic region (residues 956-1128) are critical for the pore-forming, possibly they are inserted in the membrane for further conformational changes to deliver the enzymatic domain into cytoplasm. Helix 1 is from 1046 to 1050; helix 2 is from 1055 to 1063; helix 3 is from 1073 to 1091; helix 4 is from 1117 to 1132 (Figure 3-1). These four helixes wrapping around the surface of the delivery domain. Under physiological pH, they are not hydrophobic, making TcdB protein soluble in the water solvent. While under acidic pH, the side chains of these amino acids would lose their charges and become hydrophobic. In this way these residues can act as a sensor of the environmental pH that triggers conformational changes associated with toxin domain translocation.

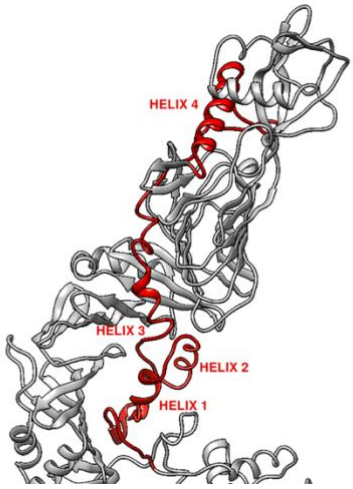


Figure 3-1 The protrusion of TcdB delivery domain.
The secondary structures are showed in ribbon style. The four helixes that are essential for pore-forming are colored in red, with their name labeled on the side.

How the conformational changes happen remains unclear. The recently published full-length TcdB crystal structure at pH 5 provides some information about a snap shot of TcdB under acidic pH (1). By comparing with its homolog model TcdA, which is another LCT toxin secreted from *C. difficile* with similar structural organization as well as toxicity mechanism, under physiological pH, the biggest difference is from residue 1024 to residue 1048. A significant movement was detected, in company with the missing structure between 1032 to 1047 residues, which form an α-helix, are now invisible, suggesting this region is highly flexible. However, because TcdB is in highly package with the nanobodies that interact on the delivery domain, CROPS domain as well as the APD domain in the crystal structure, they helped to stabilize the its structure

while limits its movement. Thus we need to find a better way to observe more TcdB states while it is in the acidic pH environment.

Cryo-EM single particle analysis is well-known for providing multiple states of the target macromolecular complexes. With fast-freezing in the liquid ethane, the particles that are in different states are locked in their conformations. By performing 3D structural classifications, we are able to capture the possible states out from the large dataset. This chapter will examine the results of cryo-EM analysis on the full-length TcdB at a condition of pH 5.

3.2. Methods

3.2.1. Protein preparation

Plasmid DNA of pHis1522 encoding wild type TcdB was transformed into *Bacillus megaterium* cells and inoculated overnight in 10 µg/mL tetracycline. The overnight culture was diluted 1:100 and inoculated to OD_{600} = 0.3 in 1 L terrific broth medium (Difco) at 200 RPM, 37 ^{o}C. 5% (w/v) xylose was added to the culture and incubated overnight at 200 RPM, 37 ^{o}C. The cell pellet was collected by centrifuge the culture at 15,000 RPM for 30 minutes. The recombinant TcdB in the pellet was purified via Ni-NTA affinity column essentially as described previously by Yang and colleagues (112). The collected pellet was lysed in PBS buffer using sonication. The Ni-NTA affinity column was washed with high-salt PBS (20 mM NaH_2PO_4, 20 mM Na_2HPO_4, 300 mM NaCl [pH 7.4]) containing 25 mM imidazole, and the bound protein was eluted using high-salt PBS containing 250 mM imidazole. The eluted protein was then loaded

onto a superdex 200 increase 10/300 GL column (GE Healthcare) and all the fractions are confirmed using SDS-PAGE.

3.2.2. Cryo-EM of TcdB at pH5

3.2.2.1. Cryo-EM sample preparation

The purified TcdB (0.3mg/mL) was pre-incubated with FZD2-CRD (purchased from R&D system) in 50mM PBS buffer at 1:1 ratio at room temperature at pH 7.4 prior to the pH change. The protein mix solution of the protein complex solution was mixed with 50mM citric acid buffer (the mixture of sodium citrate-citric acid, adjusted to pH 5) at ratio 7:3 before applying on the grid. This buffer ratio was tested in large scale by mixing 21 mL PBS buffer with 9 mL citric acid buffer (pH 5). The final pH was measured to confirm that by mixing these two buffers, a final pH of 5 is achieved. The final volume 3 uL mixture was incubated at room temperature for 30 s and then applied to C-flat 2/1 holey carbon film 300 mesh grids (Electron Microscopy Sciences) at 20˚C with 100% relative humidity and vitrified using a Vitrobot (Mark III, FEI Company, the Netherlands). The grid was then fast frozen into liquid ethane and then transferred into liquid nitrogen for further storage.

3.2.2.2. Cryo-EM data collection

The grid of TcdB-FZD2 at pH5 was imaged under the Titan Krios G3 transmission electron microscope (Thermo Fisher Scientific) operated at 300 kV. The microscope is equipped with a Gatan K3 summit direct detection camera (Gatan, Pleasanton, CA); 11,000 micrographs were collected using electron-counting mode at pixel sizes of 0.732 Å. The beam intensity was adjusted to $7e^{-}/Å^{2}/s$ on the camera. A 30-

frame movie stack was collected for each micrograph, with 0.2 seconds per frame, for a total exposure time of 6 seconds.

3.2.2.3. Cryo-EM data processing

The collected micrographs were motion corrected using MotionCorr2 (113). The dose-weighted micrographs were visually screened, and 8,434 micrographs with strong power spectra were selected for further processing. Contrast transfer functions of the micrographs and the following processing were all done in cryoSPARC (129). Particles were picked using blob-picking tool in the first round. Then the picked particles were used to generate the initial 2D classification result. This 2D classification result was used for template particle picking tool. The picked particles were scaled to a pixel size of 4.24 Å before put into final 2D classification. 664,373 clean particles were selected for a heterogeneous refinement into four classes. The particles from class that had the most reasonable structure were used for homogenous refinement and generated a density map around 4.1Å. A None-Uniform refinement was sub sequentially used to get better resolved density map and generated a 3.9 Å density. The overall resolution was assessed using the gold-standard criterion of Fourier Shell Correlation, with a cutoff at 0.143, between 2 half maps from 2 independent half-sets of data.

Multibody analysis was performed using RELION 3.0 (116) on this data set. The density map was divided into three bodies. Each of the body was done principal component analysis. A 3D classification was done on the clean particles, separating particles into four classes by applying a solvent mask only around the tip of the delivery domain when put “skip align” as one of the settings. The final result showed three states

of the delivery domain of the complex. The particles corresponding to each of the class were used for 3D refinement and three density maps were generated, representing the three states of the complex.

3.2.2.4. Model building

The 3.9 Å density map was used for TcdB-FZD2 at pH5 model building. We chose the TcdB model from previously published paper, describing the crystal structure of TcdB at pH5 with three antibodies stabilizing its structure. This model was modified based on the density map of the TcdB part and deleted the residues that are not included in the density map. The FZD2 model was used based on my previous result. The complex structure was roughly fitted in the density map in Chimera (118) and then used MDFF (119) to refine the model. Finally Phenix was used to further refine and validate the model (130).

3.2.3. Leakage assay

3.2.3.1. Liposome preparation

Liposomes were prepared as previously described, with minor modifications (131). For the structure study, 70% DSPC (Avanti), 20% cholesterol (Avanti), and 10% Ni-NTA DGPC (Avanti) were mixed and dried under a stream of dry argon, vacuum desiccated to remove residual solvents, re-suspended, with freezing and thawing, to 2 mg/ml in liposome extrusion buffer (50 mM HEPES, 150 mM NaCl, 25 mM KCl, pH 7.4) and then extruded through polycarbonate filters in a mini extruder (Avanti) for two rounds, with indicated diameters of 200 nm in the first round of 11 passes and 50 nm in the second round of 15 passes.

For the Leakage Assay, brain lipid Folch (Avanti, Cat. 131101P) from Avanti and Sigma were mixed 1:1, with 0.06% Vybrant DiD (Invitrogen, Cat. V22887). With a similar preparation process, the mixture was hydrated to 1 mg/ml in liposome extrusion buffer with 1 mM Calcein (Invitrogen) and then extruded through polycarbonate filters with the indicated diameters of 200 nm with 15 passes in a mini extruder. Extruded sample were centrifuged at 20,000 g for 30 minutes and re-suspended with liposome extrusion buffer to apply to a PD10 Desalting Column (Cytiva). Collected fractions were used as liposomes for the leakage assay.

3.2.3.2. Leakage Assay with Multicolor-Burst Analysis Spectroscopy (MC-BAS)

MC-BAS is a multi-channel extension of a single particle technique, Burst Analysis Spectroscopy (BAS) (132), which can simultaneously determine both size and stoichiometry distributions of fluorescent nanoparticles. MC-BAS can collect fluorescent signals from two different channels at the same time (133). In this way, when a particle carrying two different fluorescent dye is measured in MC-BAS, burst from these two channels are recorded at the same time point in separate panels.

To apply MC-BAS measuring the leakage of liposomes, the liposome membrane carried Vybrant DiD, while the lumen contained Calcein. Two channels, with excitation wavelength of 488 nm and 642 nm, were used to measure burst signals from membrane and lumen. For one double-labeled liposome, bursts measured from membrane and lumen were recorded at the same time point. In this way, bursts from these two channels were recorded and aligned as coincident bursts. Meanwhile, when detergent, such as Tween-20 is applied to liposomes, the entire lumen contents are expected to be released

from the liposome into the reaction buffer, due to the solubilization of liposome membrane. Since the burst from single Calcein is very small, less than 20 cps, an increase in the baseline was observed in the channel of lumen and the burst signals from two channels lose the coincidence.

3.3. Results

3.3.1. TcdB has conformational changes in the delivery domain under pH5

3.3.1.1. Overall description of TcdB conformational changes

After testing multiple conditions, TcdB was found to not aggregate in high concentration when it is in the acidic environment for short time such as 30 seconds or 1 minute. Thus, the single-particle analysis was done on TcdB at pH 5 for 1 minute to reveal the pre-pore state of TcdB. The overall density map of TcdB is at 4.1Å, with better resolved GTD and APD, where bulky side chains are observed and a less resolved delivery domain and CROPS domain, due to their flexibility. By comparing the TcdB-FZD2 complex at physiological pH (7.4) and acidic pH (5.0), it is noteworthy that the overall structure of TcdB-FZD2 becomes “expanded” and “loose”, with a ~15 Å expansion (Figure 3-2A) when measuring the distance between the tip of GTD and the tip of the delivery domain.

The most significant movement is around the rod-shape delivery domain, with a twisting movement on the beta-sheet scaffold as well as the flipping movement that is observed in CSPG4 bound state. A flexible loop between residue 944 to 958 locating at the shoulder of the GTD extends toward the delivery domain, which was not observed in physiological condition (Figure 3-2B).

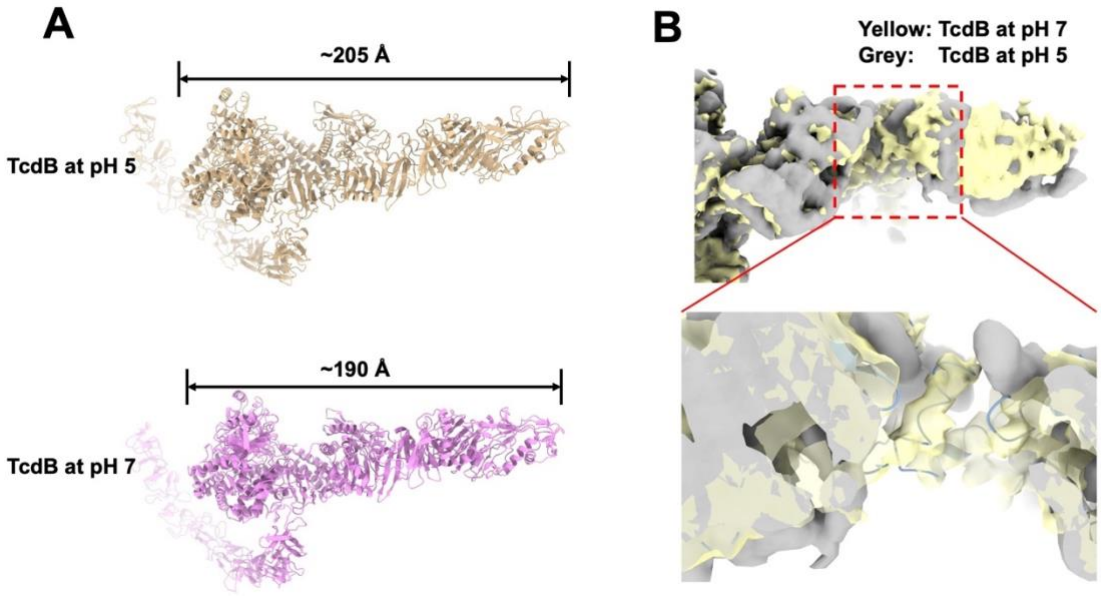


Figure 3-2 Expansion (panel A) and movement on TcdB-FZD2 (panel B) at pH5 model.
Panel A shows the model for TcdB at acidic pH (yellow) and at physiological pH (pink). The length between the delivery domain tip and the GTD tip is labeled. Panel B shows the superimpose of the TcdB densities at two pH. The density difference is highlighted in the bottom panel with a zoom in view.

Since the hydrophobic region of TcdB is critical and showed to be flexible, a 3D variability analysis was performed to reveal the movement of this domain in order to get more information. The 3D variability is an algorithm that treat each particle picture as one snapshot of the macromolecule. When considering the dataset of the input particles, it tried to separate different states and generate the trajectory of each component. In this dataset, I asked for three trajectories for calculation. In the first trajectory, movement of the CROPS domain is observed, with a missing density at the C-terminus CROPS domain. The second trajectory demonstrates that delivery domain is moving

horizontally, with a missing piece in the connection between delivery domain and APD domain (domain 1), which is involved in the "pH sensing" step as described previously (91). Movement of the "core-region", consist of GTD and APD is observed in the trajectory 3. This analysis revealed that the conformational changes of TcdB is not limited in the delivery domain. It tends to be a cooperative, fully movement all over its domains.

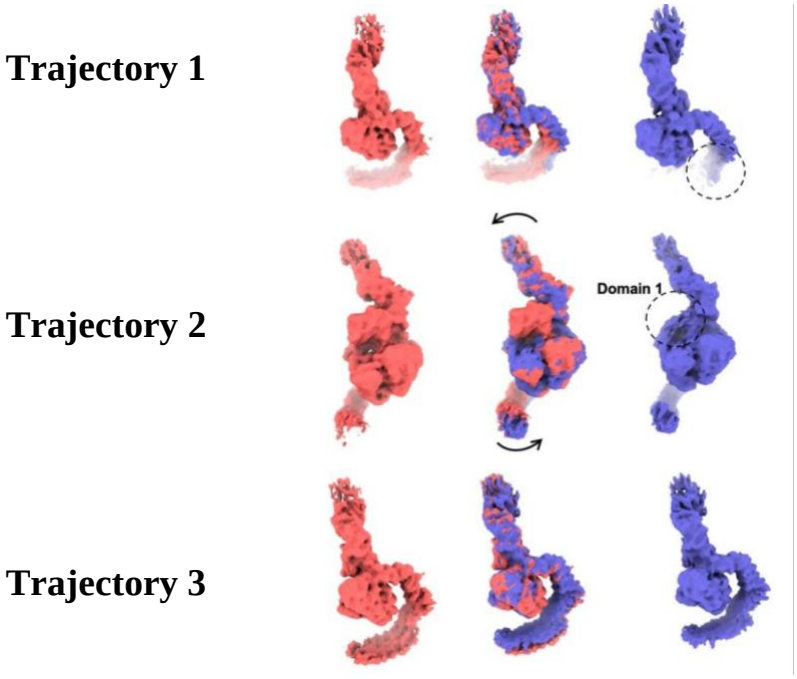


Figure 3-3 The 3D viability of TcdB.
The three trajectories are listed. The red color maps represent one extreme end of each trajectory and the bule maps represent the other extreme end of each trajectory. The missing densities are highlighted by black dash circles.

3.3.1.2. The interaction between TcdB and FZD2 is influenced by pH change

To understand the behavior of TcdB and receptors, ELISA assays on the binding-state TcdB at either pH 7.4 or pH 5 were performed. These results demonstrated that the binding of FZD2 is decrease by pH changes by ~25% while the interaction with CSPG4 is significantly reduced since the signal from CSPG4 is complete gone after pH changed (Figure 3-3). As we know CSPG4 has an extensive extracellular domain (~220 kD), with its most cell-distal N-terminal domain interacting with TcdB. Releasing TcdB from CSPG4 benefits the interaction between TcdB and the inner face of the endosome membrane. In contrast, associating with the smaller FZD2 (~15 kD) may draw TcdB into close proximate with the membrane. This makes the following step sterically possible.

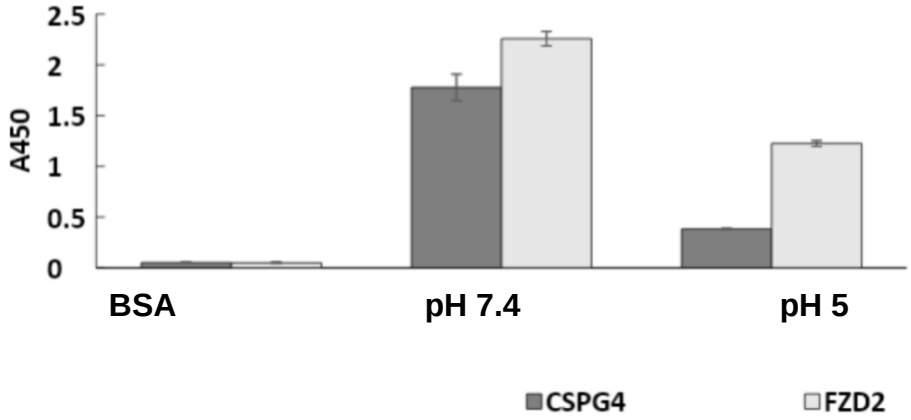


Figure 3-4 ELISA assay on the interaction between TcdB and FZD2/CSPG4 at pH7 and pH5.
The signals from interacting with CSPG4/FZD2 are showed in columns colored in deep grey/light grey.

Consistent with the ELISA result, cryo-EM density demonstrates that only ~53% of particles have well-refined FZD2 density on TcdB, suggesting flexible conformational changes happen in the connection between FZD2 and TcdB, thus alignment on FZD2 is not as easy as at pH 7.4. To reveal more details in the binding at pH5, including the interaction pattern and the conformational changes, more classification needed to be done at the interaction region locally. By applying a mask around the delivery domain, the dataset could be classified into four distinct classes (Figure 3-5). The four classes can be divided into two groups: FZD2 bound state and FZD2 dissociated state. Even in the FZD2 bound state, it is obvious that the connection between FZD2 and TcdB is very weak. When comparing the TcdB-FZD2 at neutral pH and the acidic pH, a dissociation motion is clearly observed. This indicates when pH changes, FZD2 gradually dissociate with TcdB while the conformational changes happened in TcdB.

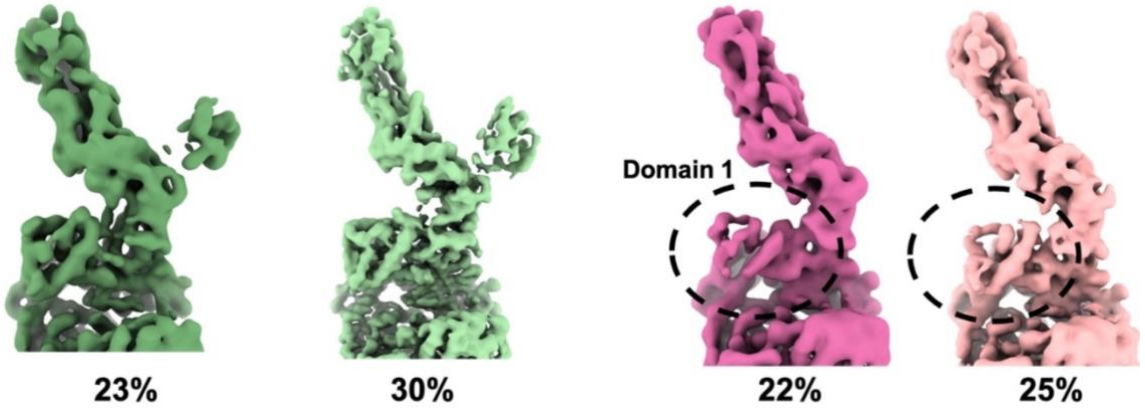


Figure 3-5 3D classification of TcdB on delivery domain.
The densities of the four classes are showed. The particle percentage that contribute to density are listed.

3.3.2. Leakage assay support that TcdB is creating holes on liposome

The liposome membrane was labeled with lipophilic carbocyanine DiD and the liposome lumen with calcein. When observing the liposome using a dual-wavelength epifluorescence microscope, the signals from both fluorophores will be detected simultaneously, as they locate on the same liposome. From the diagram, the position and the signal strength detected from DiD should consistent with the signal position and strength from calcein (Figure 3-6. A). If leakage occurs, the signal from the liposome lumen will decrease, while the signal from the membrane-bound DiD remains similar. Thus, the consistency between these two signals is lost. In other words, the coincidence between the two signals is lost. In the same time, the signal baseline from calcein will be raised comparing with the non-leakage state (134). If the leakage on the membrane causes the breakage of liposome, then both signals will not be detected. In Figure 3-6C, loss of coincidence is detected when comparing liposome signal with/without TcdB interacting. The baseline of calcein is raised. These phenomenoa indicate that there is leakage in the liposome membrane. When comparing with the positive control (adding Tween-20), the lumen signal was not drastically decreased after pore-formation (Figure 3-4B), suggesting that the leakage caused by TcdB didn't break the continuity of the liposome membrane. Thus, we can conclude that after interacting with liposome at pH 5 for 5 minutes, pore formation is happening on the liposome, and the liposome stay as intact structure in the meanwhile. However, this assay is not able to detect if the translocation has happened or not in the condition described above.

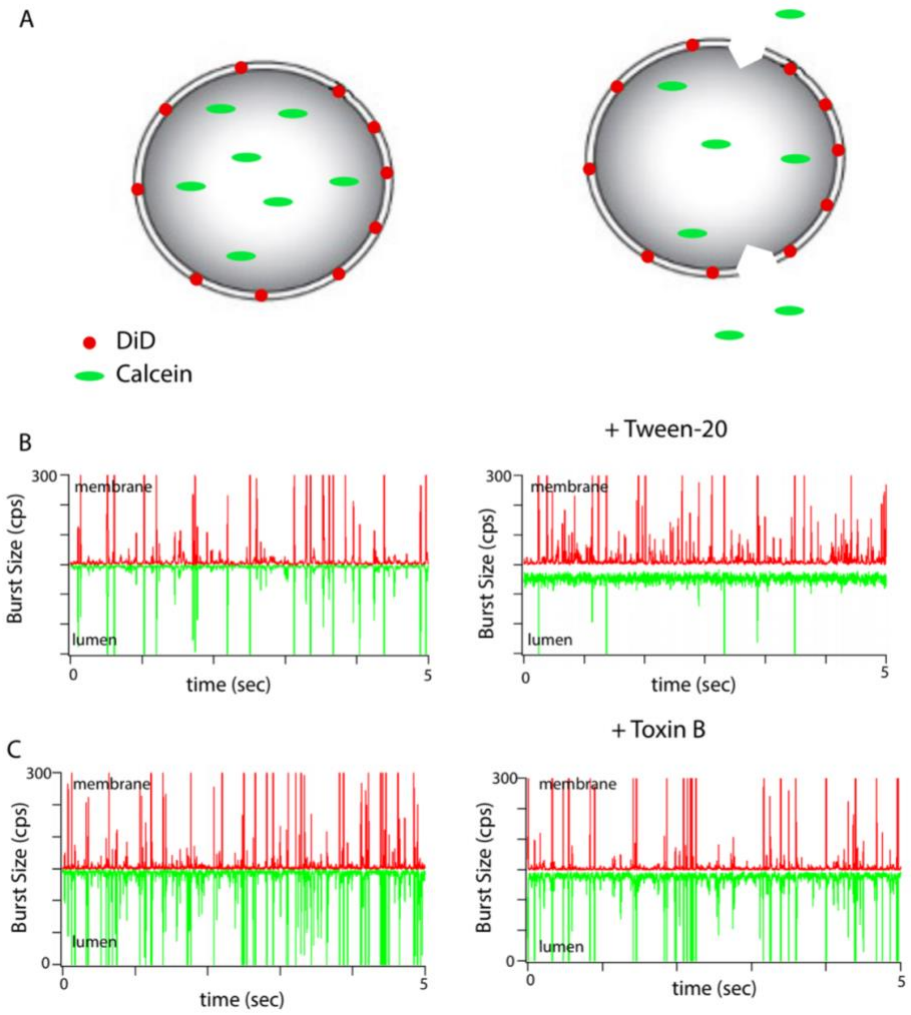


Figure 3-6 Leakage assay of TcdB on liposome.
Panel A shows the scheme of the leakage assay. Panel B shows the positive control by adding Tween-20 to the liposome. Panel C is the experimental data when adding 800uM TcdB into liposome and incubated the mixture for > 5minutes.

3.4. Discussion

The pore-forming and translocation of TcdB at acidic pH could be unique and complicated compared to other toxins that take action upon pH changes. From our cryo-

EM single particle analysis, no TcdB oligomerization was observed in our refinement, and TcdB was observed as a monomer following a 1 minute exposure to a pH 5 environment. The overall structure of TcdB is “loose” and “extended”, comparing with TcdB under acidic pH. The most significant difference is at the delivery domain, where the tip of delivery domain is getting longer and the arrangement of the beta-sheets is getting loose, suggesting a beginning of conformational change. The density of the N-terminus of the delivery domain, referred to here as “Domain 1”, is partially missing in the 3D variability analysis, indicating flexibility in this region. When pH changes, the interaction between TcdB receptor FZD2 and itself is reduced, which agrees with cryo-EM results that the density of FZD2 in cryo-EM reconstructions was significantly weaker than the density at physiological pH (~50% decrease), suggesting FZD2 might interact with TcdB in a different manner at acidic pH, possibly in a weaker and less ordered association. The leakage assay confirmed that TcdB is able to form pore on the artificial bilayer without breaking the liposome membrane. Five minutes after TcdB interacting with liposome at acidic pH is the timepoint where the pore-forming is detectable. This only conclude that there is pore formation happened on the liposome membrane but I cannot get any information about TcdB translocation. Based on previous studies, the current released from Rb^+ loaded artificial liposome lumen when TcdB creates hole on the membrane is “flickering”, which is different from other similar toxins (128), releasing constant current when the pore-forming is in action, such as anthrax toxin (83). From these result I propose a more complete picture of the TcdB conformational changes under acidic pH: When in the endosome, TcdB gradually

dissociates with its receptors as it changes its conformation. The delivery domain extends and twists away from the other domain, making the pore-forming helices exposed on the protein surface. Then the delivery domain gradually twists and bends, making these helices closer and ready to form pores on endosome membrane.

To get a complete view of TcdB pore-forming and translocation is quite difficult. The greatest obstacle is the flexibility and the potential flickering movement at the point of membrane insertion. Until now we only understand that pore-formation is happening. However, it is possible that the membrane insertion is not stable, and the relative positions of the membrane-inserting helices may also be flexible. Moreover, the most dramatic conformational changes still have not been directly observed, which would allow for firmer conclusions on the mechanism of pore formation.

To stabilize the transmembrane region of the membrane protein, liposome are used to capture the target protein and provides a suitable environment for its transmembrane region. This technique is applied on many membrane protein structural solving (135). When analyzing TcdB pore-forming and translocation, liposomes can be used to stabilize TcdB structure at acidic pH. In this work, cryo-EM data on liposome-bound TcdB was collected at pH 5 t to reveal the pore-forming state of the TcdB structure. In order to get the translocation state TcdB structure, the condition that actually generates detectable translocation state TcdB particles need to be identified. There are two positions in the GTD and APD that can be biotinylated. Preloading the liposome lumen with biotin and enzymes that are required for biotinylating is the prerequisite for the detection of pore-forming and translocation. When the TcdB is

interacting with liposome from outside at acidic pH, pore-forming and translocation is happening. If GTD and APD is translocated from outside to inside of liposome, the GTD and APD will be biotinylated. By performing western blotting against the biotin on TcdB, I am able to tell if translocation is happening or not. Signal detected in the western blotting suggest successful translocation. Structural studies will be performed after figuring out the best condition for TcdB translocation to happen.

4. SELECTION AND CHARACTERIZATION OF ULTRAHIGH POTENCY DESIGNED ANKYRIN REPEAT PROTEIN INHIBITORS OF *C. DIFFICILE* TCDB*

4.1. Introduction

Clostridium difficile is gram-positive spore-forming anaerobic bacteria. Colonization of the gut with pathogenic *C. difficile* bacteria can lead to *C. difficile* infection (CDI) with symptoms including diarrhea, pseudomembranous colitis, sepsis, multiple organ dysfunction syndrome and even death. In 2011 there were almost half a million reported cases of CDI and more than 29,000 CDI-associated deaths in the United States alone. *C. difficile* is a major nosocomial pathogen as a significant percentage (7%) patients acquire CDI after hospitalization (6). Broad-spectrum antibiotics are considered as a major culprit of CDI, as they disrupt the patients' natural gut microflora that would otherwise keep the proliferation of *C. difficile* in check. The current standard-of-care for treating CDI is to administer additional antibiotics, mainly vancomycin, metronidazole and fidaxomicin (13). Although generally effective against primary CDI, over the past decades, the rate of CDI recurrence has greatly increased due to the emergence of antibiotic-resistant and so-called hypervirulent strains (15-35% CDI recurrence in patients after cessation of antibiotic treatment) (136).

The pathology of CDI mainly stems from the two *C. difficile* secreted exotoxins, toxin A (TcdA) and toxin B (TcdB), that target small GTPases within the host cells, leading to disruption of the tight junctions and loss of colonic epithelial barrier function. Anti-TcdB monoclonal antibody, bezlotoxumab (ZINPLAVA), was approved by the FDA in 2016 for treating recurrent CDI. The CDI recurrence rate in patients receiving

antibiotics together with *i.v.* infusion of bezlotoxumab, although lower than those receiving antibiotics alone (26-28%), remains high at 15-17%. In addition, the monoclonal antibodies require mammalian cells for expression, which is much costlier compared to microbial protein expression (ZINPLAVA costs ~$3000/dose), putting a strain on the already thin health care budget.

Our long-term goal is to develop highly efficacious anti-toxin proteins as oral therapeutics for treating and/or preventing CDI. These anti-toxin proteins are based on a designed ankyrin repeat protein (DARPin), a small antibody-mimic binding scaffold that exhibits very high thermostability, resistance to protease and denaturant, and a very low immunogenicity (137). DARPins that bind to a wide range of molecules with pico- to nanomolar affinity have been identified (138). Furthermore, DARPins can be expressed at high levels in *E. coli* (multi-gram quantities per liter of culture in fermenters), enabling DARPins to be produced at potentially very low cost on a large scale.

Combining phage panning and two rounds of functional screening, a panel of dimeric DARPins yielded DARPin dimers with picomolar *in vitro* TcdB neutralization potency. The best dimeric DARPin, DLD-4, exhibited an EC_{50} of 4 pM and 20 pM against TcdB from *C. difficile* strains VPI10463 (ribotype 087) and M68 (ribotype 120), respectively, which is ~330-fold and ~33-fold more potent than the FDA-approved anti-TcdB monoclonal antibody bezlotoxumab. DARPin DLD-4 also exhibit efficacy *in vivo* in a mouse model against TcdB challenge, pointing to its potential as a next generation anti-toxin biologic for treating and/or preventing CDI. Our cryo-electron microscopy (cryo-EM) studies of the complex between TcdB and the dimer DARPin DLD-4

revealed that its two constituent DARPins, 1.4E and U3, bind to the toxin Delivery and autoprocessing domain (APD). Guided by this structural information, our subsequent ELISA studies showed that 1.4E and U3 interfere with the interaction between TcdB with its receptors CSPG4 and FZD2, respectively. Moreover, at pH 7.4, the CROPS domain in the cryo-EM structures for both apo and DLD-4-bound TcdB was found to point away from the Delivery domain, opposite to that seen in the negative-stain electron microscopy (EM) structures of TcdA and TcdB at the same pH in which the CROPS domain extends toward and kisses the Delivery domain. We believe that our cryo-EM structures of TcdB, apo and DARPin-bound, represent the native aqueous conformations that will be invaluable for future anti-toxin drug development.

4.2. Methods

4.2.1. TcdB expression and purification

Plasmid DNA encoding a 6-His tagged-TcdB was transformed into *Bacillus megaterium* cells and the recombinant TcdB was purified via Ni-NTA affinity column essentially as described previously. The column was washed with high-salt PBS (20 mM NaH_2PO_4, 20 mM Na_2HPO_4, 300 mM NaCl, pH 7.4) containing 25 mM imidazole and the bound protein was eluted using high-salt PBS containing 250 mM imidazole. Eluted protein was diluted in low salt PBS (20 mM NaH_2PO_4, 20 mM Na_2HPO_4, 10 mM NaCl, pH 7.4) to obtain a final NaCl concentration of 30 mM and loaded onto a Q HP anion exchange column (GE Healthcare). The column was washed with the same low salt PBS buffer and bound protein was eluted using a salt gradient from 10 mM to 1 M NaCl.

TcdB eluted at NaCl concentrations of ~500 mM. Protein purity was confirmed using SDS-PAGE.

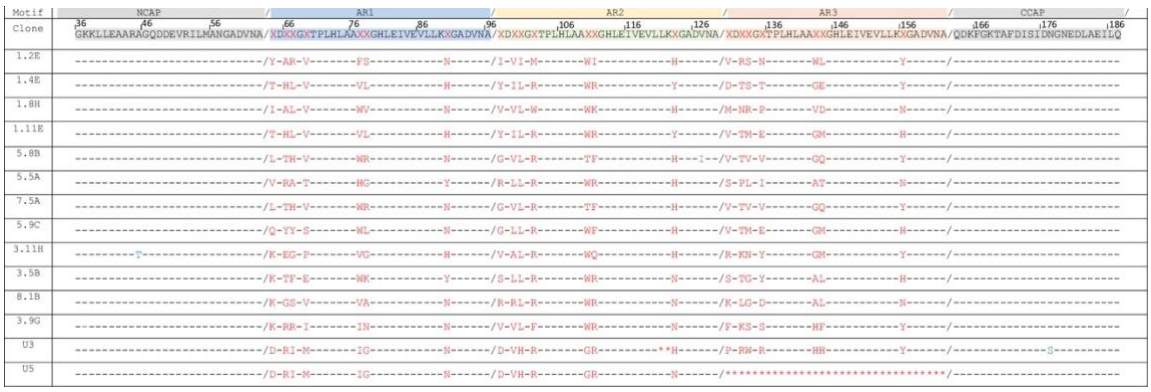
4.2.2. DARPin Library creation and phage panning

DARPins are designed based on repeat modules of natural ankyrin protein and consist of an N-terminal capping repeat (N-cap), three (N3C) internal repeats and a C-terminal capping repeat (C-cap). In a DARPin library, each internal repeat contains six randomized positions, yielding a total of 18 randomized positions in each DARPin. We prepared such a DARPin library using sequential ligation and PCR. This DARPin library was displayed as an N-terminal fusion to the bacteriophage M13 gIII minor coat protein, and after a DsbAss cotranslational translocation signal peptide. A DARPin library consisting of ~$2x10^9$ unique clones was created by transforming ~12 mL high-efficiency MC1061 electro-competent cells with ~250 µg of ligated and purified DNA.
Phage panning was carried out as described previously (139). TcdB (from *C. difficile* VPI10463) was biotinylated via EZ-Link-Sulfo NHS-LC Biotin (Pierce) and used as the target protein. Four rounds of sequential phage panning were performed. The enrichment of TcdB-binding DARPin was confirmed by phage ELISA following a published protocol (139). A plateaued level of TcdB-binding enrichment was observed after round 3 of panning, indicative of successful phage panning.

To create dimeric DARPin library, DARPin variants identified from the functional library screening were PCR amplified using Taq DNA polymerase (NEB) with two sets of primers. Set 1 uses primers Ran2-D-F and Linker-BSAi-D-R (Table 4-1) to generate DARPins with a 3’ linker sequence ((GGGGS)x3), and Set 2 uses primers

Linker-BSAi-D-F and Ran2-D-R to generate DARPins with a 5’ linker sequence. PCR products were digested with BsaI to generate sticky ends in the added linker region and ligated to form dimeric DARPins. This library was then inserted into pET26b for expression in *E. coli*.

Table 4-1 The sequence of DARPin.



4.2.3. Functional screening of TcdB-neutralizing DARPins

DARPin variants from the 3rd round of phage panning were subcloned into the pET26b vector (between NdeI and HindIII) for high-level DARPin expression. 764 individual clones of *E. coli* BL21(DE3) cells transformed with the enriched library were picked and grown in eight 96-deep well plates (1 mL/well) at 37 °C for 8-10 h. 50 µL of the overnight culture was transferred to fresh plates containing 1 mL/well LB and grown until OD_{600} ~0.6 (~3 hours) prior to the addition of isopropyl β-d-1-thiogalactopyranoside (IPTG). The culture was shaken at 400 rpm and at 37 °C for 4

hours and was harvested by centrifugation at 1700 *xg* for 20 minutes. The cell pellets were resuspended in 100 µL of PBS (1.8 mM KH_2PO_4, 10 mM Na_2HPO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4) supplemented with lysozyme (200 µg/mL), incubated at 37 °C for 30 minutes, subjected to 3 cycles of freeze-thaw between -80 °C and 37 °C, and centrifuged at 16,000 *x g* for 10 minutes. The soluble fraction was transferred to fresh 96-well deep plates and incubated at 70 °C for 20 minutes and centrifuged again, yielding highly enriched DARPin in the supernatant. The supernatant was transferred to fresh plates and stored at -80 °C until use.

Semi-purified DARPin (0.1 – 10 µL lysate) was incubated with purified TcdB in growth medium (DMEM supplemented with 10% fetal bovine serum (FBS), non-essential amino acids, penicillin (100 mg/mL) and 100 mM streptomycin) in 96-well plates for 2 hours at room temperature. After incubation, different amounts of the mixtures (0.1 – 10 µL) were added to Vero cells seeded the night before in growth medium (final TcdB concentration 132 pg/mL). The concentration of TcdB was selected so that the viability of toxin-treated Vero cells is 10-20% that of naïve Vero cells after 6 hours of toxin contact time. Cell supernatants were replaced with fresh growth medium six hours later, and the cell viability was quantified 72 hours post toxin addition using the CellTiterGlo reagent (Promega) and normalized to Vero cells treated with the equivalent amount of lysate from untransformed BL21(DE3) cells and in the absence of TcdB.

For the dimeric DARPin functional screen, the protocol was further simplified. 1504 individual clones of *E. coli* BL21(DE3) cells transformed with the dimeric

DARPin library were picked and grown in 16 deep 96-well plates (1 mL/well) at 37 °C and 400 rpm in LB overnight. The next day, the culture was harvested by centrifugation at 1700 *xg* for 20 minutes, and each of the cell pellet was resuspended in 200 µL PBS supplemented with lysozyme (200 µg/mL), incubated at 37 °C for 30 minutes, subjected to 1 cycle of freeze-thaw between -80 °C and 37 °C, and incubated at 70 °C for 20 minutes. An equivalent amount of 0.2 µL of the lysate (after serial dilution) was added to Vero cells together with TcdB toxin (10 pg/mL). 72 h later, the cell viability was quantified by the CellTiterGlo assay and normalized to that of naïve Vero cells.

4.2.4. Protein expression and purification

E. coli BL21(DE3) cells transformed with DARPin in pET26b were cultured overnight at 37 °C in auto-induction media (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 20 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, 0.6% glycerol, 0.1% glucose, 0.08% lactose) supplemented with 50 µg/mL kanamycin. Cells were lysed by sonication. The lysate was clarified by centrifugation at 16,000 *xg* for 10 minutes, and the soluble lysate was filtered through a 0.45 µm PES membrane and loaded onto a gravity Ni-NTA agarose column. The column was washed with PBS containing 15 mM imidazole and the bound proteins were eluted using PBS containing 150 mM imidazole. Protein purity was determined using SDS-PAGE.

DNA encoding bezlotoxumab was synthesized. The anti-TcdB mAb was expressed in CHO cells and purified following standard protocols in the Feng laboratory.

CSPG4 extraceullar domain (410-560) was expressed based on the protocol described in Section 2.

4.2.5. ELISA

MaxiSorp immunoplates (Nunc) were coated with 4 µg/mL TcdB overnight at 4 °C. The next day the wells were washed and blocked with PBSTB buffer (PBS containing 0.1% Tween-20 and 2% BSA) before being incubated with serially diluted IMAC-purified DARPins (containing a myc-tag at the N-terminus), CSPG4-EC-GFP, FZD2-EC (R&D systems). After incubation, wells were washed 4 times with PBST. Bound DARPins were detected using mouse anti-c-myc antibody (1 µg/mL, Invitrogen and HRP-conjugated anti-mouse antibody (0.8 µg/mL, JacksonImmunoResearch). Bound CSPG4-EC-GFP was detected using rabbit anti-GFP antibody (0.05 µg/mL, Proteintech) and HRP-conjugated anti-rabbit antibody (0.8 µg/mL, Santa Cruz Biotechnology). HRP-conjugated goat anti-human antibody (0.025 µg/mL, JacksonImmunoResearch) was used to detect bound FZD2-EC. The color development agent was 3,3',5,5'-tetramethylbenzidine (TMB).

4.2.6. *In vitro* TcdB neutralization assay

Vero cells ($1.5x10^3$ cells/well) in growth medium were seeded in 96-well plates. The next day, serial dilutions of IMAC-purified DARPins were added to the appropriate wells followed by the addition of TcdB (final concentration 5 pg/mL or 2.5 pg/mL). The concentration of TcdB was selected so that the viability of toxin-treated cells is 10-20% that of naïve Vero cells after 72 hours of toxin contact time. The plates were incubated at 37 °C/5% CO_2 for 72 hours and after which the cell viability was quantified using the CellTiter-Glo cell viability kit (Promega).

4.2.7. *In vivo* TcdB Neutralization Activity of DARPins

Six to eight week-old CD1 mice were purchased from Harlan Laboratories (MD, USA). All mice were housed in dedicated pathogen-free facilities in groups of 5 mice per cage under the same conditions. Food, water, bedding, and cages were autoclaved. All procedures involving mice were conducted under protocols approved by the Institutional Animal Care and Use Committees at the University of Maryland. DNA encoding bezlotoxumab was synthesized and the antibody was purified from the supernatant of transiently transfected CHO cells. DLD-4 (2.5 mg/kg or 0.25 mg/kg), or bezlotoxumab (10 mg/kg) was mixed with TcdB (1.5 μg/kg) in PBS and incubated at room temperature for 1h before being injected intraperitoneally (i.p.) into CD1 mice in the appropriate treatment groups (5-10 mice /group). The control group was i.p. injected with TcdB alone in PBS. Mouse survival was monitored for 4 days until the termination of the experiments and data were analyzed by Kaplan–Meier survival analysis with Log rank test of significance.

4.2.8. Electron-microscopy sample preparation

TcdB and DLD-4 were mixed at 1:1 molar ratio (with the final concentration of the complex at 800 nM) and incubated in PBS buffer at pH 7.4 for 30 minutes at room temperature. 3 μL of the complex was applied to C-flat 1.2/1.3 holey carbon film 400 mesh grids at 20 ^{o}C with 100% relative humidity and vitrified using a Vitrobot (Mark III, FEI company, Netherlands). Apo-state TcdB was prepared in the same buffer and vitrified using the same condition but without the incubation at room temperature.

4.2.9. Cryo-EM data collection

The complex of TcdB and DLD-4 was imaged under both the FEI Tecnai F20 electron microscope (FEI company, Netherlands) and a JEOL JEM3200FSC electron microscope (JEOL, Japan), with their respective field emission gun operated at 200 kV and 300 kV. Both microscopes are equipped with a Gatan K2 summit direct detection camera (Gatan, Pleasanton, CA). 2,255 (from Tecnai F20) and 1,647 (from JEM3200FSC) micrographs were collected using macro mode or manual mode of SerialEM[60] in the super-resolution electron counting mode. A nominal magnification of 25,000X (on Tecnai F20) and 30,000X (on JEM3200FSC) were used, yielding subpixel size of 0.75 Å and 0.6 Å, respectively. The beam intensity was adjusted to 5 $e^{-}/Å^2/s$ on the camera. A 33-frame movie stack was collected for each picture, with 0.2 s per frame, for a total exposure time of 6.6 s. For data collected on the JEM3200FSC, an in-column energy filter was used with a slit width of 29 eV.

Apo-state TcdB was imaged similarly but under the FEI Tecnai F20 electron microscope only. Image data was collected under nominal magnification of both 25,000 X and 29,000 X, yielding subpixel size of 0.75 Å and 0.62 Å, respectively.

4.2.10. Image processing

Movie stacks, collected on both microscopes for the complex of TcdB and DLD-4, were first binned by 2 to yield pixel size of 1.5 Å and 1.2 Å. Using Unblur (140), these stacks were aligned and summed to generate a set from frames 1–33. These sum images were visually screened, and 1,272 and 1,173 of each set with strong power spectra were selected for further processing. Contrast transfer functions of the

micrographs were estimated using CTFFind4 (141). Batchboxer in EMAN (142) was first used to automatically pick all the particles from 1,272 sum images with a box size of 216 pixel^2, or particles from 1,173 sum images with a box size of 280 pixel^2, yielding 331,516 particles and 234,642 particles, respectively. These particles were scaled to a common pixel size of 3 Å. The automatically picked particles were then screened for high-contrast particles for four rounds of the reference-free 2D classification in RELION (116). 45,806 and 38,614 clean particles were selected and combined for 3D classification, separating particles into four classes. Since density maps from these classes had similar conformations except for the CROPS domain, all the clean particles were used for 3D refinement to generate a final density map of TcdB-DLD-4 complex at 9 Å resolution. The same particles were also used for 3D refinement of the core region with a solvent mask that masked out the CROPs region, generating a density map of the TcdB-core-DLD-4 complex at 8.3 Å resolution. The overall resolution was assessed using the gold-standard criterion of Fourier Shell Correlation, with a cutoff at 0.143, between two half maps from two independent half-sets of data. Local resolutions were estimated using Blocres (143). The apo TcdB structure was similarly processed.

4.2.11. Model building

The homology models for the core region of the TcdB (residues 1-1799), the U3 and 1.4E modules of the DLD-4 were built in Swiss-model. The linker between U3 and 1.4E was manually built in UCSF chimera (118). These models, along with the crystal structure of the N-terminal half of the CORPS domain in TcdB (residues 1834-2101, pdb code: 4NP4) were combined in UCSF chimera and refined into the cryo-EM density map

using Molecular Dynamics Flexible Fitting to generate the complex structure of TcdB and DLD-4. I-TASSER was used to predict the secondary structure for the region of residues 1800-1834.

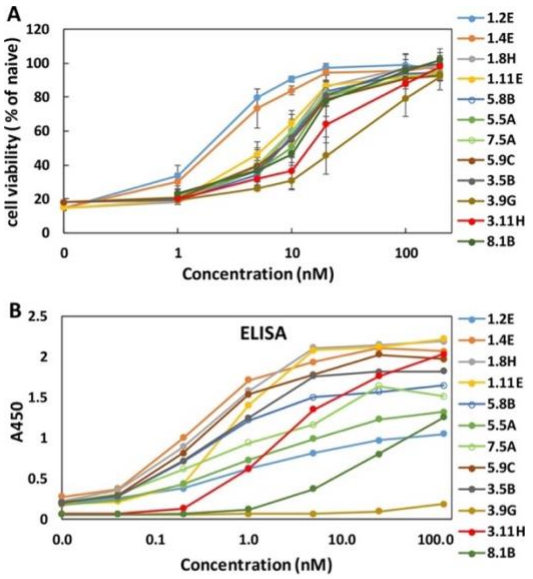
4.3. Results

4.3.1. Dimeric DARPins inhibits TcdB toxicity in a more effective way

4.3.1.1. Engineering monomeric TcdB-neutralizing DARPins

A library of approximately $2x10^9$ DARPin variants was constructed via sequential PCR and ligation essentially as described previously. Biotinylated TcdB (from *C. difficile* strain VPI10463) was used as the target protein to enrich DARPins that could bind the toxin via four rounds of phage panning. DARPin variants from the 3rd round of phage panning were subcloned into the pET26b vector and transformed into *E. coli* BL21(DE3) cells for high-level DARPin expression and functional screening for those with toxin-neutralization ability. About 40% of the clones (299 clones) were able to rescue Vero cells viability from TcdB toxicity by >50%. The top 40 hits were sequenced and of which 12 unique clones were identified and further characterized. Most clones exhibited EC_{50} values of ~10 nM, and the 2 best clones, 1.2E and 1.4E, showed EC_{50} of 2.4 nM and 3 nM, respectively. The relative affinity of each of the top 9 DARPins for TcdB was assessed by ELISA (Figure 4-1.B), and was found to not directly correlate with their *in vitro* neutralization potency. For example, although 1.2E and 1.4E exhibit similar TcdB-neutralization ability, 1.4E appeared to be among the strongest toxin binders while 1.2E was the weakest binder. The discrepancy between binding affinity and neutralization potency likely stems from the different epitopes

engaged by the different DARPins. It is conceivable that a DARPin binding to a region critical for the toxin activity would exhibit better toxin-neutralization potency than another DARPin binding to a region of less importance even if the first DARPin binds the toxin with weaker affinity than the second DARPin.



C

DARPin	EC_{50} (nM)	DARPin	EC_{50} (nM)
1.2E	2.4 ± 0.5	7.5A	10.2 ± 3.1
1.4E	3 ± 0.7	5.9C	10.2 ± 0.9
1.8H	9.1 ± 3.3	3.11H	23.5 ± 12.4
1.11E	7.5 ± 1.2	3.5B	13.3 ± 6.9
5.8B	10.4 ± 2.3	8.1B	14.4 ± 6.6
5.5A	11.3 ± 3.0	3.9G	42.7 ± 21.2

Figure 4-1 Screen monomeric DARPin.

4.3.1.2. Engineering dimeric TcdB-neutralizing DARPins

Fusion of multiple DARPins have been reported to significantly improve the target-binding affinity via avidity effects (144). We hypothesized that fusion of two

DARPins binding to non-overlapping epitopes on the toxin should lead to greatly improved binding affinity that increases toxin-neutralization potency. A combinatorial library of DARPin dimers was created by joining individual monomeric DARPins (12 total) via a flexible linker (GGGGSx3). A total of 1504 individual clones were screened using the Vero cell assay and 12 hits were identified. Of which, 10 were determined to be unique clones. The *in vitro* neutralization potencies of these 10 DARPins and their relative TcdB binding affinities were shown in Figure 4-2. The best DARPin dimer, DLD-4, exhibited a toxin-neutralization EC_{50} of 4 ± 1 pM, which is ~600-fold lower than the best monomer DARPin, 1.2E (EC_{50} 2.4 ± 0.5 nM). Similar to that observed for the DARPin monomers, the relative TcdB binding affinities of these DARPin dimers do not directly correlate with their neutralization potency. Sequencing results also revealed that many of these dimers contain a new DARPin, U3, that is not present in the original 12 monomers. U5 is identical to U3 except that it lacks the third randomized ankyrin repeat domain. U3 alone has weak, but detectable, toxin-neutralization (Figure 4-3) and toxin-binding abilities (Figure 4-4) and is likely a minor constituent present in the initial pool used to create the dimer DARPin library.

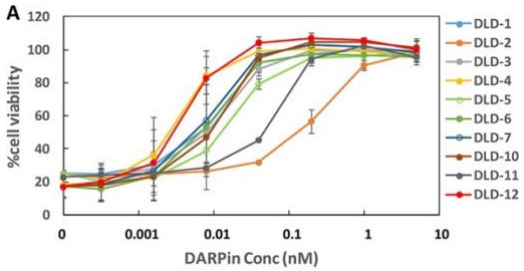
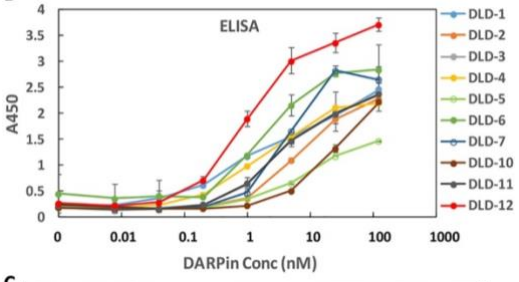


Figure 4-2 Dimeric DARPin screening.

B



C

DARPin dimer	N-DARPin	C-DARPin	EC_{50} (nM)
DLD-1	5.5A	7.5A	0.43 ± 0.375
DLD-2	1.8H	1.2E	1.04 ± 0.594
DLD-3	U3	5.9C	0.0515 ± 0.054
DLD-4	U3	1.4E	0.004 ± 0.001
DLD-5	U5	5.8B	0.031 ± 0.013
DLD-6	U3	3.5B	0.0125 ± 0.011
DLD-7	U3	5.5A	0.014 ± 0.014
DLD-10	U3	5.8B	0.005 ± 0.001
DLD-11	U3	1.2E	0.290 ± 0.109
DLD-12	U3	1.11E	0.008 ± 0.005

Figure 4-2 Continued.

To understand the reason for the dramatic improvement in activity, we further characterized the five best dimer DARPins. We first compared the *in vitro* TcdB-neutralization potency of the DARPin dimers with their constituent monomers (Figure 4-3). All these DARPin dimers significantly out-performed their constituent monomers and the combination of both monomers, indicative of synergistic activity among the constituent monomer DARPins. We further compared the relative binding affinity of DARPin dimers and monomers using ELISA. As shown in Figure 4-4, except for DLD-6, the combination of the two constituent monomer DARPins lead to additive increase of ELISA signal, suggesting that these constituent monomers bind non-overlapping epitopes on the toxin. Dimeric DARPins showed further enhanced binding/ELISA signal, indicative of avidity effect between the constituent DARPin monomers. An

exception is DLD-6, which appeared to bind the toxin weaker than its constituent DARPin 3.5B at concentrations <25 nM. In addition, unlike other dimeric DARPins where the combination of constituent DARPin monomers leads to an additive increase of binding signal, no increase in binding signal was observed with both constituent DARPins (3.5B + U3) than with 3.5B alone. This result is somewhat unexpected, considering that the toxin-neutralization potency of DLD-6 is > 100 fold higher than that of 3.5B (EC_{50}: 12.5 pM for DLD-6 vs. 13.3 nM for 3.5B). We posit that one possible cause of this discrepancy is that TcdB protein, when immobilized on ELISA plate, exhibits less flexibility, which may hinder the simultaneous interaction by both U3 and 3.5B.

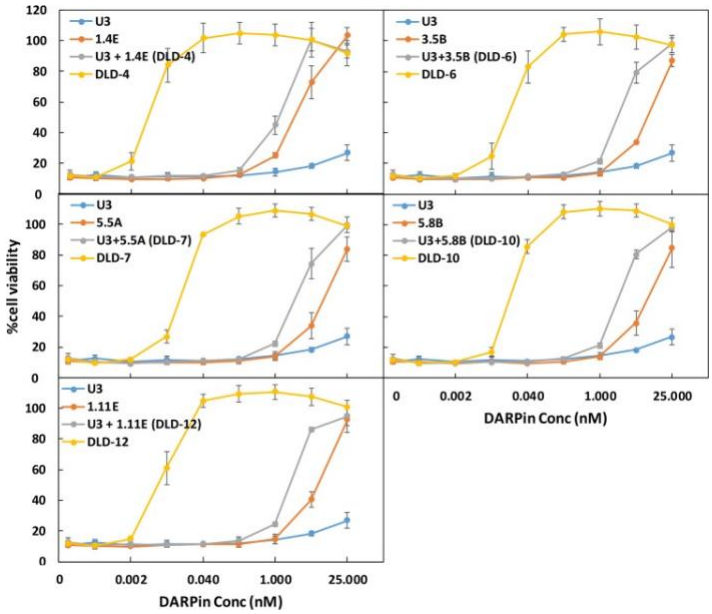


Figure 4-3 DARPin dimers exhibit superior toxin-neutralization potency relative to the constituent monomers.

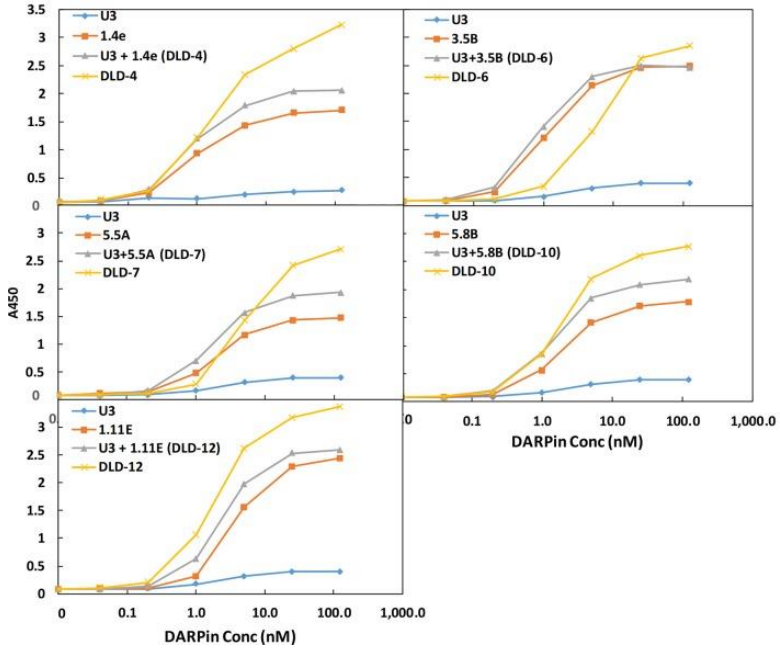


Figure 4-4 DARPin dimers demonstrate avidity in TcdB binding.

4.3.1.3. Characterization of DARPin TcdB-neutralization potency

All of our protein engineering work was conducted using TcdB from the laboratory strain of *C. difficile* VPI10463 (ribotype 087). However, there is a significant amount of amino acid sequence heterogeneity between different strains of *C. difficile* and thus a need to develop broadly neutralizing DARPins. As a first step to address this need, we evaluated the activity of selected DARPin dimers (DLD-4, -7, -11, -12) against TcdB from three different strains of *C. difficile*: VPI10463, M68 (ribotype 012) and UK1 (ribotype 027). All DARPins were effective against toxins from VPI10463 and M68 (Figure 4-5). The best DARPin, DLD-4, was found to be ~330- and ~33-fold more efficacious than bezlotoxumab at inhibiting TcdB from VPI 10463 and M68,

respectively (Figure 4-5B). However, no neutralization activity was observed for any of the DARPins against TcdB from UK1 strain, which belongs to the hypervirulent 027 ribotype. Bezlotoxumab also showed significantly weaker, albeit detectable, neutralization activity against this toxin (EC_{50} >25 nM). We are currently repeating our *in vitro* engineering using TcdB from the UK1 strain of *C. difficile* as the target protein in order to identify DARPins efficacious against this toxin.

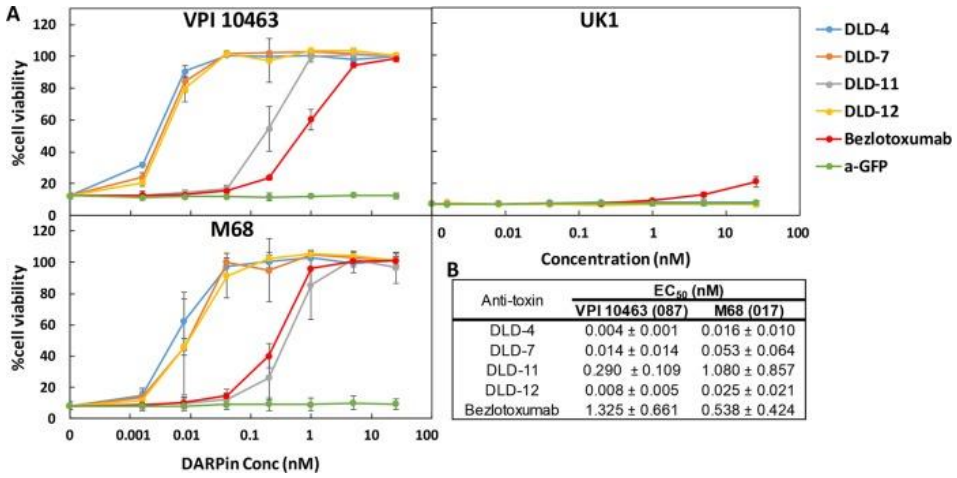


Figure 4-5 DARPin dimers offer superior protection to Vero cells against the toxicity of TcdB from *C. difficile* strains VPI 10463 (ribotype 087) and M68 (ribotype 017).

We further evaluated the ability of the most potent DARPin, DLD-4, to protect mice from systemic toxin challenge *in vivo*. A lethal dose of TcdB (1.5 µg/kg) was mixed with DLD-4 (0.25 or 2.5 mg/kg), bezlotoxumab (10 mg/kg) or PBS and then injected intraperitoneally (i.p.) into CD1 mice (5-10 mice/group). Bezlotoxumab is the

monoclonal antibody that was approved by the FDA in 2016 for treating recurrent *C. difficile* infection and was used here as a control. 40% of the mice survived after injection with the mixture containing the toxin mixed with 2.5 mg/kg DLD-4 (p=0.04), while no measurable protection was observed in mice receiving the same dose of toxin mixed with bezlotoxumab (10 mg/kg) or DLD-4 (0.25 mg/kg), as all mice died within 30 hours post injection (Figure 4-6). In previous studies, when a much lower toxin dose was used (25 ng/mouse or ~1 µg/kg), i.p. injection of bezlotoxumab (10mg/kg) lead to 38% mice (3 out of 8) surviving 72 h post toxin injection[29]. These results demonstrated that the DLD-4 possesses significantly higher toxin-neutralization ability than bezlotoxumab.

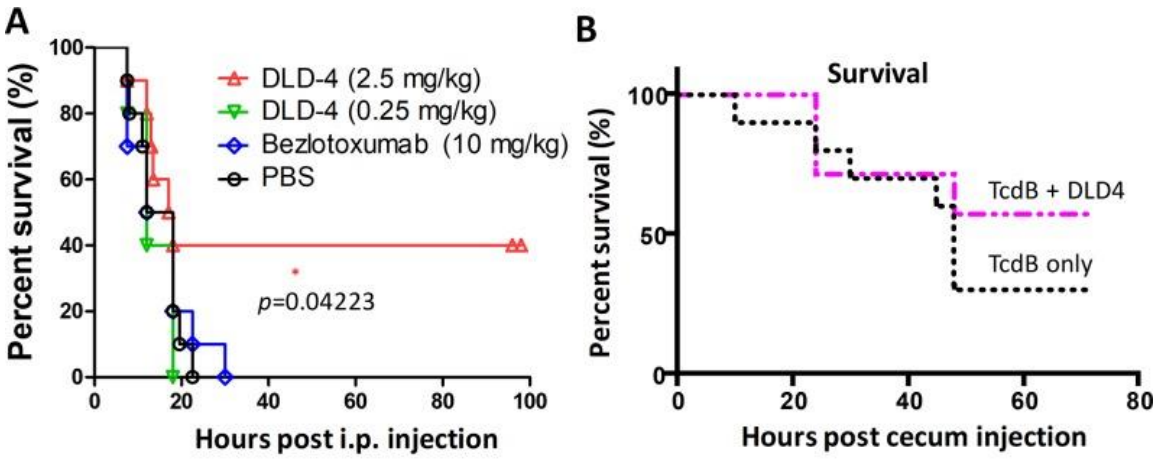


Figure 4-6 In vivo studies on DARPin neutralization ability.

4.3.2. Dimeric DARPin interacts with TcdB on the receptor binding pockets

4.3.2.1. Cryo-EM structure of the full-length TcdB and its interaction with DARPin DLD-4

To understand how DLD-4 neutralizes TcdB, single-particle cryo-electron microscopy (cryo-EM) was used to determine the complex structure of TcdB and DLD-4.

The overall structure of full-length TcdB is very similar to TcdA, consisting of four functional domains: the glucosyltransferase domain (GTD), the autoprocessing domain (APD), the Delivery domain and CROPS domain. The Delivery domain undergoes conformational changes triggered by the low pH environment in the late endosome to form pores, translocating the GTD across endosome membrane into the host cytosol after ADP-catalyzed intramolecular cleavage reaction. Once inside the cytosol, GTD inactivates small GTPases such as RhoA, Rac1 and Cdc42 by glucosylation, causing a loss of actin polymerization and cytoskeletal changes and disruption of the colonic epithelial junctions.

The two constituent DARPins of DLD-4, U3 and 1.4E, bind around the middle and the C-terminal end of the Delivery domain, respectively (Figure 4-7A-C). The 1.4E also interacts with the C-terminal region of the APD. The resolved protein secondary structures allowed us to model each domain of the TcdB and DLD-4 into the cryo-EM density (Figure 4-7D). Notably, the density of the 15-residue long peptide linker between the C-terminus of U3 and the N-terminus of 1.4E was visible, enabling the assignment of each DARPin into its corresponding density (Figure 4-7E). The local

resolutions of the cryo-EM density are not uniform throughout the complex, with lower resolutions around U3 of DLD-4 and the C-terminal tip of the CROPS domain. These are consistent with the observation that U3 DARPin binds relatively weakly to TcdB and the tip of the CROPS domain exhibits larger flexibility, both of which contribute to their lower local resolutions in single-particle cryo-EM reconstructions.

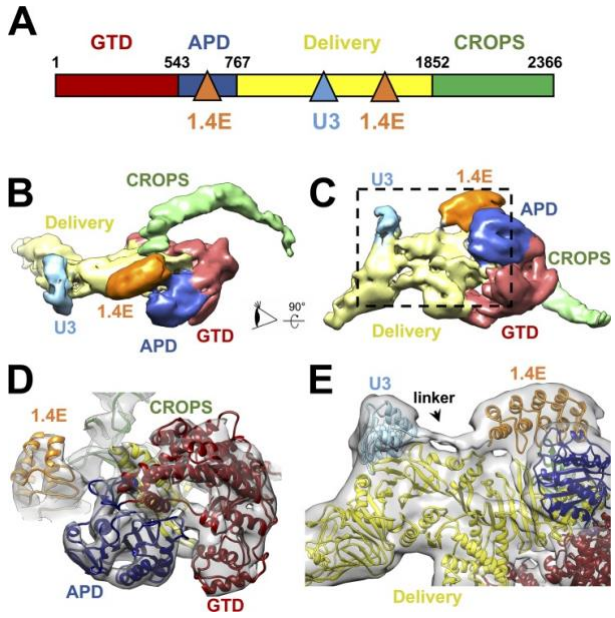


Figure 4-7 Cryo-EM structure of the TcdB–DLD-4 complex.
A) Domain organization of TcdB and its interaction with the 2 modules of the DLD-4 DARPin. The binding sites on TcdB for DLD-4 are indicated by the blue and orange triangles for the U3 and 1.4E modules, respectively. (B) Density map of the TcdB–DLD-4 complex with the functional domains and DARPin modules in different colors. (C) The same map but rotated 90° as indicated by the arrow. (D) The model fitted into the map as viewed from the eye cartoon labeled in Panel B. The density map is iso-surfaced at a threshold of 9 σ. (E) Zoom-in of the region labeled within the dashed black box in Panel C to show the linker between DARPins U3 and 1.4E. The density map is iso-surfaced at a threshold of 3.8 σ. APD, autoprocessing domain; CROPS, combined repetitive oligopeptides; cryo-EM, cryo-electron microscopy; DARPin, designed ankyrin repeat protein; GTD, glucosyltransferase domain; TcdB, *C. difficile* toxin B.

3D classification of all the cryo-EM particles of TcdB and DLD-4 revealed multiple conformations of the CROPS domain (Figure 4-8), indicating a continuum of conformational variations at the C-terminal tip of the CROPS domain. Importantly, in all observed conformations of both DLD-4-bound and apo TcdB particles, the CROPS domain projects away from the Delivery domain. Such a conformation of the CROPS domain is strikingly different from that reported in the negative-stain electron microscopy (EM) structure of the full-length apo TcdA and TcdB at the same pH (pH 7.4), in which the CROPS domain extends toward and 'kisses' the Delivery domain. To resolve whether the 180° shift of the CROPS domain results from the difference between cryo-EM and negative-staining EM specimen preparations, we calculated the two-dimensional (2D) class-averages of the negatively stained apo TcdB particles at pH 7.4. In ~80% of the negatively stained TcdB particles, the CROPS domain extends toward the Delivery domain (Figure 4-8B), similar to that seen in the negatively stained apo TcdA and TcdB. Only ~20% of the apo TcdB particles have the CROPS domain projecting away from the Delivery domain similar to that observed in cryo-EM TcdB structure. Unlike cryo-EM, in which the protein samples are preserved in a vitreous state with water molecules in and surrounding the specimen, negative-stain EM inevitably results in dehydration and flattening of the biological specimens which may result in distortion of the molecule's conformation. Thus, we believe that the TcdB conformation in which the CROPS domain projects away from, rather than extending toward, the Delivery domain, represents the native aqueous conformation of TcdB at pH 7.4.

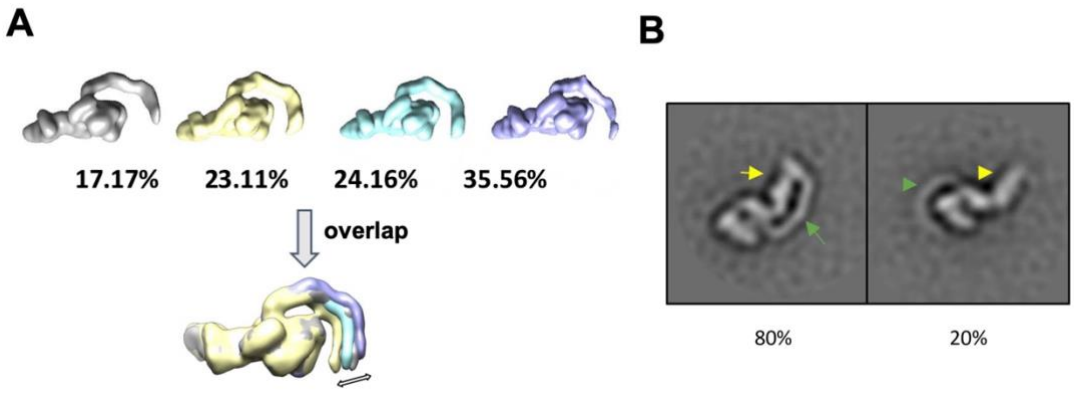


Figure 4-8 3D classification result and the negative stain result of TcdB and DLD-4 complex.

Examination of the interface between TcdB and DLD-4 revealed detailed binding modes between the toxin and the two constituent DARPins, U3 and 1.4E, which are separated by ~30Å and orientated perpendicular to each other (Figure 4-9A). The two DARPins are connected by a 15-residue linker (GGGGSx3), which lies along the side of the Delivery domain of TcdB. Each DARPin consists of 3 designed ankyrin repeat (AR) modules sandwiched between the N- and C-terminal capping ARs. Each designed AR and the C-terminal capping AR consist of a b-turn loop followed by two antiparallel a-helices, resulting in three variable and one fix loops on each DARPin. These four loops are labeled as Loop 1-4 in Figure 4-9. As anticipated, both DARPins have their variable loop regions contacting the TcdB and their helical scaffold regions exposed to the solvent. U3 grips onto the middle of the Delivery domain (a β-sheet from residues 1461-1510 and a loop from residues 1595-1603, Figure 4-9B) while 1.4E sits between the C-

terminus of the Delivery domain and the N-terminus of the CROPS domain (residues 1800-1834), and interacts with a loop from the APD (residues 747-751) (Figure 4-9C). The sequence of TcdB from 1800 to 1834 lacks a high-resolution structural reference. However, both the shape of the EM density and the secondary structure prediction based on the protein sequence indicated this region to be a β-sheet (grey density in Figure 4-9C).

The cryo-EM structure revealed several protein residues carrying opposite charges at the binding interface between DLD-4 and TcdB. For example, the U3 has Arg 66 on Loop 1 and Lys 163 on Loop 4 that are within 5-Å distances to Glu 1468 and Asp 1501 from the TcdB Delivery domain, respectively (Figure 4-9B). In the UK1 strain of TcdB, the negatively charged Glu 1468 and Asp 1501 are occupied by the positively charged Lys and polar Asn, respectively. These changes in electrostatic interactions may explain the weakened affinity of U3 toward the TcdB from the UK1 strain.

For DARPin 1.4E, the loops 1 and 2 span over a β-sheet region (residues 1749 - 1767) in the TcdB Delivery domain, while loops 3 and 4 interact with part of CROPS (residues 1800-1834) and APD (residues 747-751), respectively. Glu 144 and Lys 168 on loop 3 and loop 4 of 1.4E are within 5-Å distances to the oppositely charged Lys 1846 in the CROPS domain and Glu 749 in the APD, respectively (Figure 4-9C). Given these observations, charge interaction also likely contributed to the binding of 1.4E to TcdB. It is worth noting that this region (residues 1753-1851) was reported to be highly variable between TcdB from different ribotypes and neutralizing epitopes within this region was found to be less accessible in TcdB of ribotype 027 (e.g. UK1) than that of

ribotype 012. No detectable binding between 1.4E and UK1 strain of TcdB was observed in our ELISA assay. Thus, the lack of neutralization activity of 1.4E against TcdB from UK1 may be partially due to the occlusion of the neutralization epitope on this toxin.

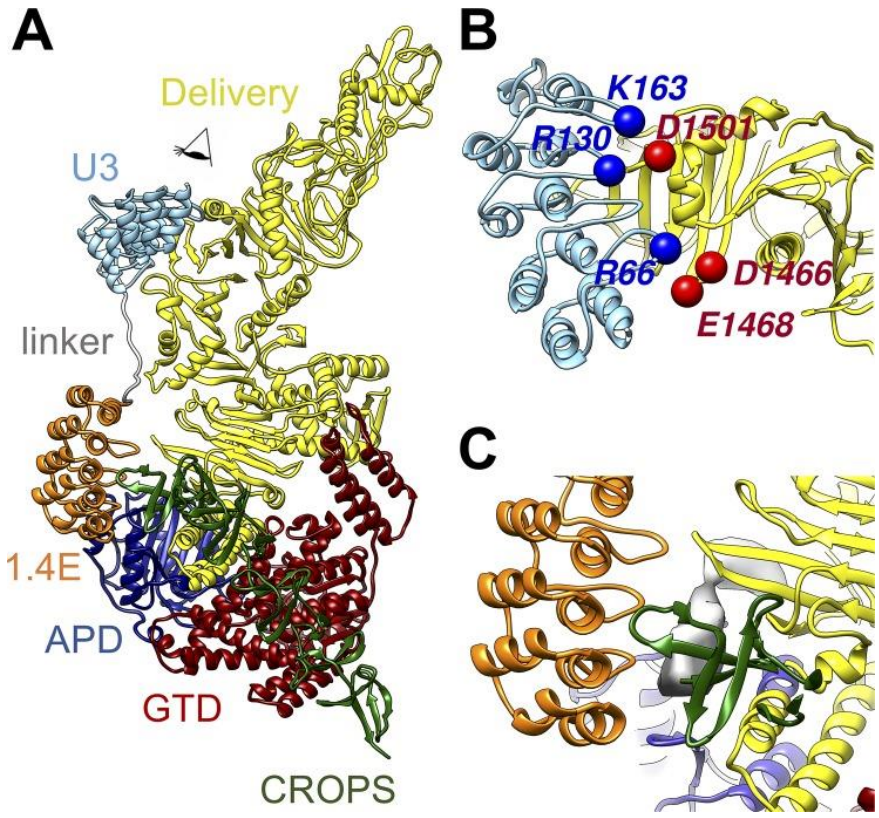


Figure 4-9 Interactions between TcdB and DLD-4.
(A) Overall model of the complex. (B) Zoom-in view to show the interactions between U3 DARPin and TcdB as viewed in the direction labeled by an eye cartoon in Panel A. (C) Zoom-in view to show the interactions between the DARPin 1.4E and TcdB at the region as labeled by red dashed box in Panel A. APD, autoprocessing domain; CROPS, combined repetitive oligopeptides; DARPin, designed ankyrin repeat protein; GTD, glucosyltransferase domain; TcdB, *C. difficile* toxin B.

4.3.2.2. Mechanism of TcdB neutralization by DARPin DLD-4

Both TcdA and TcdB enter cells via receptor-mediated endocytosis. The CROPS domain is historically thought to be the sole receptor-binding domain. However, three cell-surface receptors for TcdB have been recently been reported: chondroitin sulfate proteoglycan 4(CSPG4), poliovirus receptor like 3 (PVRL3 or NECTIN3), and members of the Frizzled protein family (FZD1, FZD2, and FZD3). Our best TcdB-neutralizing DARPin DLD-4 consists of DARPin 1.4E and U3. The Cryo-EM study revealed that 1.4 E interacts with regions between residues 747-751 and 1800-1834, while U3 interacts with regions between 1461-1510 and 1595-1603 (Figure 4-8A). Since these regions over overlap with that of TcdB receptor CSPG4 (between residues 1756-1852, designated D97), we postulated that DLD-4 may neutralize the toxin by blocking the interaction between the toxin and at least one of its receptors.

We first determined whether DLD-4 would compete with CSPG4 for binding to TcdB using competitive ELISA. A GFP-tagged extracellular domain of CSPG4 (CSPG4-EC-GFP) was expressed in HEK 293F cells and purified. For the ELISA experiments, MaxiSorp 96-well plates were first coated with purified TcdB, then incubated with CSPG4-EC-GFP (1 nM) in the presence or absence of DLD-4 or its constituent DARPins (250 nM) for 1 h at room temperature. After thorough washing, the bound CSPG4-EC-GFP was detected using anti-GFP antibody. As shown in Figure 4-9A, DLD-4 and its constituent 1.4E were able to significantly reduce the binding signal from CSPG4-EC-GFP with DLD-4 producing a more pronounced signal reduction than

1.4E. No significant reduction in binding signal was observed for U3. This result indicates that the 1.4E DARPin interferes with the toxin-CSPG4 interaction.

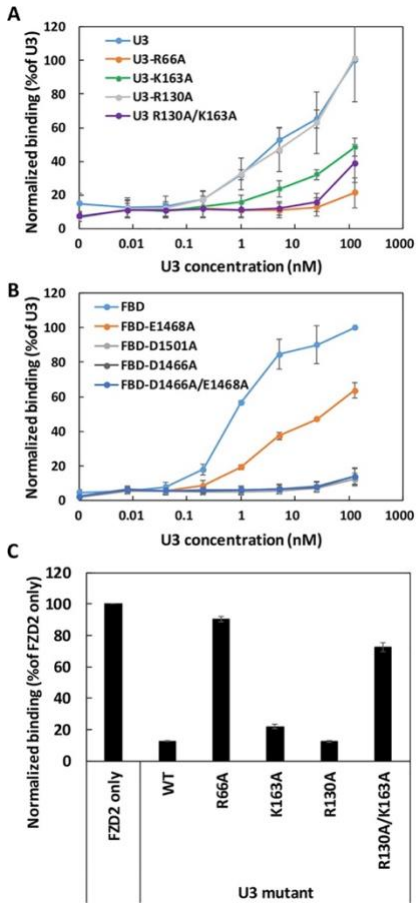


Figure 4-10 Charge contribution of binding between U3 and TcdB$_{VPI}$.
Relative binding of U3 to different immobilized mutants of FBD (A) or different U3 mutants to immobilized TcdBVPI (B) was determined using ELISA. In both cases, serially diluted WT or mutant U3 proteins were added to microtiter plates coated with 4 µg/mL of the TcdBVPI (A) or FBD variants (B). Results are the average of 3 independent experiments, and the error bars represent the standard deviation. (C) Relative binding of FZD2-Fc to immobilized TcdBVPI. The ELISA plates were coated with TcdBVPI followed by treatment with 1 nM of FZD2-Fc alone or in combination with 250 nM of WT or different U3 mutants. The error bars represent mean deviation from 2 independent experiments. ELISA, enzyme-linked immunosorbent assay; FBD, frizzled binding domain; FZD2-Fc, Fc-tagged extracellular domain of FZD2; TcdB, C. difficile toxin B; WT, wild type.

We carried out a similar competition ELISA experiment to determine whether DLD-4 and FZD2 bind to overlapping epitopes on TcdB. After toxin immobilization, the plate was incubated with FZD2-Fc (1 nM) in the presence or absence of DLD-4 or its constituent DARPins (250 nM). As shown in Figure 4-9B, significant reduction of binding signal was observed for DLD-4 and U3 while no reduction of FZD2-binding signal was observed for 1.4E. This data indicates the U3 competes with FZD2 for binding to TcdB. Since U3 interacts with TcdB between residues 1461-1510 and 1595-1603, our result essentially maps the binding footprint of FZD2 for the first time. No signal reduction was observed for the extracellular domain of PVRL3 (PVRL3-EC) using the same assay, suggesting that DLD-4 does not interfere with the interaction between PVRL3 and TcdB. Altogether, these studies led us to conclude that the high toxin-neutralization potency of DLD-4 stems from its ability to simultaneously interfere with the ability of TcdB to interact with two of its cellular receptors CSPG4 and FZD2.

4.4. Discussion

C. difficile infection (CDI) represents a serious public health problem with more than $6 billion in annual treatment-associated costs. CDI often occurs post broad-spectrum antibiotics treatment, which disrupts the patients' natural gut microflora that would normally prevent the colonization of *C. difficile*. Consequently, antibiotics are non-ideal therapeutics against CDI. In addition, the rate of CDI recurrence post cessation of antibiotic treatment has greatly increased in the past decades, pointing to a great need for non-antibiotic-based CDI therapeutics. Direct toxin-neutralization by monoclonal antibodies have emerged as a promising therapy against CDI, with bezlotoxumab

(ZINPLAVA), a fully human monoclonal antibody against *C. difficile* toxin TcdB, recently being approved by the FDA for treatment of recurrent CDI. The market value of bezlotoxumab is predicted to reach over US$212 million by 2020. However, even with bezlotoxumab, the CDI recurrence rate remains high at 15-17%. Furthermore, bezlotoxumab shows significantly reduced neutralization potency against toxins from hypervirulent strains of *C. difficile* ribotype 027 and 078.

In this study, we aimed to engineer a highly efficacious microbial-expression compatible antibody surrogate protein, designed ankyrin repeat protein (DARPin), for the neutralization of *C. difficile* toxin TcdB. DARPins represent a versatile class of binding proteins that have been engineered to bind diverse targets with up to picomolar affinity. Furthermore, DARPin is also amenable to high yield production by microbial recombinant expression hosts. Our long-term goal is to create DARPin-based oral therapeutics against CDI infection. Combining phage-panning and functional screening, we identified 12 DARPins that protected Vero cells against TcdB-induced cytopathic effect with nanomolar EC_{50} values. A secondary functional screening of dimeric DARPins yielded 10 dimers with >100-fold improved toxin-neutralization potency compared to the monomers. The best dimer DARPin, DLD-4, inhibited TcdB from VPI 10463 (ribotype 087) and M68 (ribotype 017) with EC_{50} values of 4 and 16 pM respectively, which are ~330-fold and ~33-fold more potent than the FDA-approved bezlotoxumab. Furthermore, DLD-4 effectively protected mice against lethal TcdB toxin challenge *in vivo*.

Cryo-EM studies revealed that the two constituent DARPins of DLD-4, U3 and 1.4E, bind to two different regions on the TcdB that overlaps with those associated by TcdB receptor CSPG4 (residues 1756-1852) and FZD2 (residues 1501-1834), respectively. Subsequently competitive ELISA studies confirmed that 1.4E and U3 competes with CSPG4 and FZD2, respectively, for binding to TcdB. No competition between DLD-4 and the third TcdB receptor Pvrl3 was observed. Thus, the ultra-high neutralization potency of DLD-4 likely stems from its ability to simultaneously block the interaction between TcdB and two of its receptors. In addition, it is worth noting that the current study also, for the first time, pin points the binding epitopes of FZD2 to the footprint of U3 (residues 1501-1510 and 1595-1603).

Moreover, our cryo-EM studies revealed a novel conformation of TcdB at pH 7.4 in which the CROPS domain points away from the Delivery domain. This is in contrast to what is seen in negative-stain EM structure of TcdA, as well as the major conformation of the negatively stained apo-TcdB. Given that unlike negative-stain EM, the cryo-EM specimen preparation suffers no structural artifacts from specimen preparation, we believe that our cryo-EM structure likely represents the native conformation of TcdB in aqueous environment at pH 7.4.

In summary, we report the engineering of a panel of DARPins with superior toxin neutralization potency against *C. difficile* toxin TcdB than the FDA-approved anti-TcdB monoclonal antibody bezlotoxumab. These DARPin-based highly potent anti-toxins possess great potential as therapeutics to treat and/or prevent CDI. Cryo-EM structural studies, for the first time, revealed a novel and native conformation of the full-

length toxin at pH 7.4, encompassing the CROPS domain and provided important structural insight on the toxin-neutralization mechanism-of-action of DLD-4. These results should enable better design and engineering of inhibitors against TcdB.

5. CONCLUSIONS

5.1. Receptor binding of TcdB

In this dissertation, I examined in detail the interaction between the *Clostridium difficile* toxin TcdB and the FZD2/CSPG4 receptors of the host intestinal epithelium. This is the first time the structures of full-length TcdB bound to its receptors have been determined. The interactions between TcdB and FZD2/CSPG4 are fundamentally different. The interaction between TcdB and FZD2 is primarily hydrophobic and confirms a published crystal structure (89). Because a homolog model for CSPG4 had been lacking, this is the first time that the structure of a CSPG4 fragment and its glycosylation site have been visualized. The interaction interface between TcdB and CSPG4 was identified, although the exact pattern of the interaction remains unknown. I not only described the interaction sites of TcdB with CSPG4 but also found that the autoprocessing domain (APD) of TcdB is involved in receptor binding. Even though the combined repetitive oligopeptides (CROPS) domain of TcdB is proposed to interact non-specifically with the glycan (145), the glycan moiety of the CSPG4 fragment (residues 410-560) seems not to be close to TcdB, suggesting that glycosylation of this CSPG4 fragment is not involved in the binding.

By comparing the density maps of TcdB in its FZD2 and CSPG4 bound states, I can tell that the delivery domain of TcdB is relatively "stable" when interacting with FZD2, comparing with CSPG4 binding state. This conclusion is supported by my 3D classification focused on the delivery domain region. Nearly all particles converge into one class for the delivery domain in the FZD2-bound state of TcdB, demonstrating that

all particles adopt a similar domain organization. The reason could be that FZD2 stabilizes the delivery domain and stop it from flipping around. When TcdB interacts with CSPG4, the TcdB delivery domain adopts three different conformations classified as CSPG4 binds to the N-terminus of the CROPS domain.

The delivery domain goes through conformational changes, including bending and twisting motions, when the pH decreases. This acidification-induced movement of the delivery domain of TcdB bound to CSPG4 will therefore be less restricted than when TcdB is bound to FZD2. A study demonstrating that cells expressing CSPG4 in addition to FZD2 are more susceptible to TcdB intoxication than cells that only express FZD2 (81) supports this hypothesis.

Cryo-EM gives a more dynamic and natural state of macromolecules than X-ray crystallography because the molecules are fast-frozen in vitrified ice. Protein particles that are in different states can be classified and resolved if they are present in different discrete states. Here, I provide not only a detailed analysis of the interaction of TcdB with its receptors but also more dynamic views of TcdB when it interacts with its receptors. These results provide more information about the mechanism of TcdB uptake and may suggest future approaches to the development of CDI treatment.

5.1.1. Why does TcdB utilize a different receptor-binding mechanism than TcdA?

It is known that TcdB and TcdA utilize different regions of the proteins to interact with different receptors. Only the grooves formed by LRs and SRs in the CROPS domain of TcdA have been shown to interact with its receptor such as sucrase-isomaltase and gp96, whereas TcdB can enter the host cell by interaction with multiple

different receptors. Because TcdA and TcdB have a similar structural organization of their CROPS domains, one might have thought that they should share a similar entry mechanism. When observing the detailed amino acids in the grooves of LRs and SRs in TcdB, reversed charges of critical binding residues are found when comparing the primary sequence with TcdA. For example, the carbohydrate from the cell membrane is proposed to interact with the acidic side chain of glutamic acid residue 2532 in TcdA, which is replaced by a lysine residue at the equivalent position in TcdB. This flip-of-charge makes it impossible for the amino acids in the TcdB CROPS domain grooves to interact with glycosylated membrane protein like TcdA does. Thus, the surface of the TcdB CROPS domain is not sufficient for interaction with glycan.

After interacting with their receptors, TcdA/TcdB go through well-defined endocytosis pathways. TcdB is incorporated through the clathrin-mediated pathway, whereas TcdA enters the cell through the caveolae-mediated pathway (146) (147). These two pathways are both utilized by enteritic viruses and toxins. For example, the Ebola and SARS viruses go through the clathrin-mediated pathway, while Tetanus toxin and *E.coli* shiga toxin go through the caveolae-mediated pathway (148). In the caveolae-mediated pathway, the caveolae aggregate underneath the receptor and form vesicles in the cytoplasm. However, not all of these vesicles merge with the early endosome, which means that some portion of the TcdA will not get the chance to complete its action. Developing multiple receptors results in more efficient incorporation. TcdB is reported to have a greater cellular toxicity than TcdA, and its use of multiple receptors may contribute to this effect. Even though there are seven binding pockets for carbohydrates

in the TcdA CROPS domain, the binding affinity between each binding pocket and its receptor is relatively low (~$5X10^2$ M^{-1}) (149), whereas TcdB interacts with its receptors with a much higher affinity (~20nM).

5.1.2. Does TcdB incorporate into cells using one receptor or multiple receptors?

FZD2 has a relatively small extra cellular domain (~30kD), whereas the extracellular domain of CSPG4 is much bigger and heavily decorated by glycans. Binding with these two receptors significantly increases the rate of endocytosis. However, the size difference between CSPG4 and FZD2 makes it hard to imagine that TcdB is able to interact with both receptors at the same time. Previous studies demonstrated that TcdB binds to CSPG4 andFZD2 independently, and the toxicity is additive (81). This means that TcdB probably interacts with either CSPG4 or FZD2 prior to endocytosis. Based on our ELISA assay results for binding of CSPG4 and FZD2 with TcdB at neutral and acidic pH, I found that when TcdB is internalized through binding with CSPG4, it dissociates from the receptor more quickly, probably because conformational changes of the TcdB delivery domain are less restricted. TcdB dissociates from FZD2 relatively slowly, and it binds to the delivery domain, two circumstances that presumably slow down pore formation and translocation of TcdB out of the late endosome into the cytosol.

5.1.3. Future questions needed to be resolved

Even though my structures revealed the interactions between TcdB and its two receptors, FZD2 and CSPG4, the interaction between TcdB and its third potential receptor, PVRL3, remains unclear. In LaFrance et al. (79), the authors propose that

PVRL3 is a TcdB receptor because: 1) A direct interaction was found between purified TcdB and PVRL3. 2) Anti-PVRL3 antibody blocks the cytotoxicity induced by TcdB. However, a knock-down of PVRL3 levels did not reduce the potency of TcdB in causing rounding of Hela cells. Also, expressing PVRL3 in a CSPG4-knockdown cell line did not restore TcdB toxicity, whereas complementation with FZD2 made the CSPG4-knockdown susceptible to TcdB. Thus, although PVRL3 may be involved in the mechanism of TcdB uptake, its exact function needs to be investigated more carefully. One possibility is that PVRL3 interacts with TcdB prior to its internalization. However, this interaction does not lead to endocytosis but rather blocks some other cellular activities related to PVRL3, such as the regulation of cell adhesion. Another possibility is that interaction with PVRL3 increases the local concentration of TcdB, thereby making it easier to contact the receptors that actually trigger endocytosis. To test this possibility, one could generate different truncated variants of TcdB to identify in the region that interacts with PVRL3.

5.2. Pore formation and translocation

Pore formation and translocation out of the late endosome are the least understood steps in the uptake of TcdB. It is impossible to separate the pore-formation step from the translocation step as they happen nearly simultaneously (128). Our cryo-EM results provide some intermediate states of TcdB at acidic pH that provide some idea of how these processes occur. Based on a 3D variability analysis, a continuous movement, rather than discrete states, is observed, indicating that the movement of TcdB is very fast and involves a flip. Direct observation of the different intermediate

states is quite difficult based on our preliminary results. Another unique aspect of this process is that the pore formed by TcdB in the endosome membrane is not present continuously. This conclusion is supported by the observation that the leakage of Rb^+ from inside liposomes is pulsatile (128). Also, no oligomerization of TcdB is observed, either when it is in citric acid buffer or bound to the liposome membrane. This observation is totally different from anthrax toxin, which forms heptamer when pH decreases.

Based on the results presented in Chapter III, I propose a model for pore formation and translocation of TcdB. TcdB is in a flexible state, with its delivery domain flipping around and its CROPS domain moving. When the pH decreases, drastic conformational changes occur, the whole particle expands, and the delivery domain tip extends forward. The helices (residues 1018-1112) that wrap around the delivery domain also become extended. The short helices formed by residues 1043-1050 and 1054-1065 are flexible and insert into the endosome membrane. The loop comprised of residues 935 to 949, which are involved in the translocation are flipped. These changes accompany large conformational changes in domain 1 in the beginning of the delivery domain, which are proposed to be related to the translocation.

Even with so many movements observed, it is hard to imagine that TcdB is able to move the GTD and APD (~90kD) from the lumen of the endosome to the cytosol just by inserting several helices into the membrane. Based on the results from the leakage assay, we know that TcdB is able to form a pore in the liposome membrane within five minutes after acidification. Thus, I should be able to utilize cryo-EM to solve a rough

structure for TcdB within the liposome membrane. If a large domain reorganization occurs, I should be able to detect it, and with the liposome present, the TcdB structure will be highly stabled. The rate of insertion of TcdB into the membrane remains unknown. Optimizing conditions, such as liposome composition and the buffer conditions, is necessary to find the most efficient insertion rate of TcdB. In order to detect the insertion rate of TcdB, I will add protease in the citric acid buffer that contains TcdB/liposome mixture. If the translocation is happening, the TcdB GTD/APD will be protected from protease digestion as it is flipped in the liposome.

5.2.1. The role of the CROPS domain in TcdB pore formation

Chen et. al. (1) compared the particle lengths of TcdB at physiological pH and acidic pH using small-angle X-ray scattering. They found that, at acidic pH, TcdB adopts an "open" state in which the CROPS domain points away from the delivery domain. At physiological pH, TcdB adopts a "closed" state in which the CROPS domain is aligned with the delivery domain. Our cryo-EM structures show only the "open" state under both pH conditions. This could be because TcdB stays in the "closed" state for a very brief time. Knowing the "closed" structure and comparing it with the "open" state, we might be able find out more about the function of the TcdB CROPS domain and its role in regulation of the conformational change. This will require collecting more cryo-EM data to capture this transition state and the application of model-orientated classification to identify the rare particles that exist in the "closed" state.

5.2.2. Is TcdB able to influence the morphology of the endosome?

Detectable signal mismatch and an increase in the baseline are observed in the leakage assay, suggesting that: 1) There are not many holes formed in the liposome membrane that will disrupt its intactness after pH decreases. We can still observe strong signals from calcein that remains in the liposome lumen, and only some of the calcein signal does not co-localize with the membrane that contains lipophilic carbocyanine DiD 2). There are signs that liposomes are getting much closer when liposomes are incubated with TcdB at pH5 for 5 minutes. This was shown by comparing the signal from DiD on the liposome membrane. When TcdB was added at pH5, the signal from liposome membrane became stronger, and with less TcdB only a medium strength signal existed, suggesting that some of the medium-sized liposomes become larger somehow after treatment with TcdB. This was demonstrated by our cryo-EM data of TcdB added to liposomes at pH5. The neighboring liposomes are not fusing but connected together by TcdB particles.

Whether this phenomenon is significant to the toxicity of TcdB is not known. When performing the leakage assay, I used a high concentration of the liposome and TcdB mixture to replicate the conditions used for cryo-EM single-particle data collection. Thus, the liposomes might have aggregated because of the exposure of the hydrophobic patch of TcdB on the liposome exterior.

5.3. Concluding remarks

As mentioned in the main text, TcdB secreted from *C. difficile* is a highly flexible and dynamic protein that is suitable to be studied using cryo-EM. By collecting

large data sets of different states of TcdB, I was able to visualize the dynamic movements that accompany receptor binding and the conformational changes that occur when the pH in the endosome lumen decreases. The two receptors for TcdB interact with different regions of TcdB. FZD2 binds to the middle of the delivery domain, forming a hydrophobic patch with a protrusion of the delivery domain. The interaction with FZD2 stabilizes TcdB, especially the delivery domain. pH change has no effect on the binding affinity of FZD2 for TcdB, which might position the delivery domain close to the membrane, ready for insertion. Whether stabilizing the TcdB delivery domain under physiological pH benefits the following conformational change remains unknown. A more complete picture of the conformational changes in the delivery domain need to be resolved in order to understand the stabilizing effect of FZD2. I have collected thousands of micrographs of His-tagged TcdB bound to liposomes. The liposome provides a hydrophobic environment for the TcdB pore-forming region and stabilizes the local structure.

A model of the CSPG4 fragment (residues 410-560) that binds TcdB is presented in this dissertation. CSPG4 is a highly glycosylated membrane protein that is important for the regulation of many pathways, including the cell cycle and cell morphology. When a protein is highly glycosylated, it is impossible to solve its structure with crystallography. A high resolution density map was derived in this work and provides a good model of the fragment, with the glycan moieties visible. Based on our structure, the binding of CSPG4 and TcdB depends on a hydrophobic interaction at the N-terminus of the CROPS domain and a charge-charge interaction with the APD. A 3D classification

performed on the TcdB-CSPG4 data revealed a flipping movement at the tip of the delivery domain. This movement is not seen with TcdB bound to FZD2. Acidification significantly decreases the binding affinity of TcdB for CSPG4. Because the N-terminal domain of CSPG4 is far from the membrane, TcdB bound to CSPG4 is not positioned to insert into endosomal membrane. Thus, dissociation from the receptor is necessary for the TcdB to translocate through the endosome membrane.

Single-particle cryo-EM was also performed on FZD2-bound TcdB at acidic pH. After 1 min at acidic pH, conformational changes are observed, but they are not drastic enough to allow pore formation. Thus, the structure I determined must represent an intermediate pre-pore state of TcdB. Although aggregations were found when TcdB at a high concentration (> 0.1mg/mL) was incubated in citric acid buffer at pH5, no obvious oligomerization was found during the single-particle analysis, indicating that the pore-forming mechanism of TcdB is distinct from that of other AB toxins.

It has been proposed that TcdB pore-forming is a transient behavior, meaning that TcdB may not stay inserted in the membrane or create a permanent hole that allows trafficking of small molecules from the lumen to the cytosol. This hypothesis was supported by a leakage assay conducted with liposomes. When TcdB was attached to the exterior surface artificial liposomes saturated with Rb^+ in their lumen, lowering the pH induced a flow of Rb+ from the inside to the outside. This leakage was transient, suggesting that TcdB does not form a long-lived channel after pore formation. Based on this result, we propose that TcdB may adopt multiple states as it inserts and then retracts from the membrane. We saw great flexibility of the CROPS and delivery domains at

acid pH, together with subtle movements of the GTD and the APD. These observations all suggest that TcdB is in an extreme unstable state at acidic pH.

6. REFERENCE

1. *Intestinal flora in newborn infants with a description of a new pathogenic anaerobe, Bacillus difficilis.* **Hall IC, O'Toole E.** 1935, Am J Dis Child, pp. 390-402.

2. *Laboratory Maintenance of Clostridium difficile.* **Dineen, Joseph A. Sorg Sean S.** 2009, Current protocols in microbiology , pp. 9A.1.1-9A.1.10.

3. *Faecal toxin and severity of antibiotic-associated pseudomembranous colitis.* **Burdon DW, George RH, Mogg GA, Arabi Y, Thompson H, Johnson M, Alexander-Williams J, Keighley MR.** 1981, J Clin Pathol., pp. 34(5):548-51.

4. *Occurrence of toxin-producing Clostridium difficile in antibiotic-associated diarrhea in Sweden.* **Aronsson B, Möllby R, Nord CE.** 1981, Med Microbiol Immunol., pp. 170(1):27-35.

5. *Prevention of beta-lactam-associated diarrhea by Saccharomyces boulardii compared with placebo.* . **McFarland LV, Surawicz CM, Greenberg RN, Elmer GW, Moyer KA, Melcher SA.** 1995, American Journal of Gastroenterology, pp. 90(3):439-48.

6. *Clostridium difficile infection in the elderly: an update on management.* **Nicolau, Tomefa E Asempa and David P.** 2017, Clin Interv Aging, pp. 12: 1799–1809.

7. *Readmissions following hospitalizations with Clostridium difficile infections.* **Elixhauser A, Steiner C, Gould C.** 2009, Healthcare Cost and Utilization Project,.

8. *Burden of Clostridium difficile on the Healthcare System.* **Olsen, Erik R. Dubberke and Margaret A.** 2012, Clin Infect Dis, pp. S88–S92.

9. *Relapse versus reinfection: recurrent Clostridium difficile infection following treatment with fidaxomicin or vancomycin.* **Figueroa I, Johnson S, Sambol SP, Goldstein EJ, Citron DM, Gerding DN.** 2012, Clin Infect Dis. , pp. 55 Suppl 2():S104-9.

10. *Clostridium difficile infection.* **Stanley JD, Bartlett JG, Dart BW 4th, Ashcraft JH.** 2013, Curr Probl Surg., pp. 50(7):302-37.

11. *The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome.* **Sebaihia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R.** 2006, Nat. Genet., pp. 779–786.

12. *Treatment of first recurrence of Clostridium difficile infection: fidaxomicin versus vancomycin.* **Cornely, O. A., Miller, M. A., Louie, T. J., Crook, D. W., and Gorbach, S. L.** 2012, Clin. Infect. Dis., pp. 55, S154–S161.

13. *Vegetative Clostridium difficile survives in room air on moist surfaces and in gastric contents with reduced acidity: a potential mechanism to explain the association between proton pump inhibitors and C. difficile-associated diarrhea?* . **Jump RL, Pultz MJ, Donskey CJ.** 2007, Antimicrob Agents Chemother. , pp. 51(8):2883-7.

14. *Clostridium difficile: its disease and toxins.* **Lyerly DM, Krivan HC, Wilkins TD.** 1988, Clin Microbiol Rev., pp. 1(1):1-18.

15. *Clostridium difficile colitis.* **Kelly CP, Pothoulakis C, LaMont JT.** 1994, N Engl J Med., pp. 330(4):257-62.

16. *Bile salts and glycine as cogerminants for Clostridium difficile spores.* **Sorg, J. A., and Sonenshein, A. L.** 2008, J. Bacteriol, pp. 190, 2505–2512.

17. *Recent progress in Bacillus subtilis sporulation.* **Higgins, D., and Dworkin, J.** 2012, FEMS Microbiol. Rev., pp. 36, 131–148.

18. *Accessory gene regulator-1 locus is essential for virulence and pathogenesis of Clostridium difficile.* **Darkoh, C., Odo, C., and DuPont, H. L.** 2016, Mbio, p. 7:e01237.

19. *Initiation of sporulation in Clostridium difficile: a twist on the classic model.* **Edwards, A. N., and McBride, S. M.** 2014, FEMS Microbiol. Lett. , pp. 358, 110–118. .

20. *Updates on Clostridium difficile spore biology.* **Gil, F., Lagos-Moraga, S., Calderon-Romero, P., Pizarro-Guajardo, M., and Paredes-Sabja, D.** 2017, Anaerobe , pp. 45, 3–9. .

21. *Characterization of relA and codY mutants of Listeria monocytogenes: identification of the CodY regulon and its role in virulence.* **Bennett, H. J., Pearce, D. M., Glenn, S., Taylor, C. M., Kuhn, M., Sonenshein, A. L.** 2007, Mol. Microbiol., pp. 63, 1453–1467.

22. *CcpA-mediated repression of Clostridium difficile toxin gene expression.* **Antunes, A., Martin-Verstraete, I., and Dupuy, B.** 2011, Mol. Microbiol. , pp. 79, 882–899.

23. *Global transcriptional control by glucose and carbon regulator CcpA in Clostridium difficile.* **Antunes, A., Camiade, E., Monot, M., Courtois, E., Barbut, F., Sernova, N. V., et al.** 2012, Nucleic Acids Res., pp. 40, 10701–10718.

24. *Global analysis of the sporulation pathway of Clostridium difficile.* **Fimlaid, K. A., Bond, J. P., Schutz, K. C., Putnam, E. E., Leung, J. M., Lawley, T. D., et al.** 2013, PLoS Genet., p. 9:e1003660.

25. *Identification of a novel lipoprotein regulator of Clostridium difficile spore germination.* **Fimlaid, K. A., Jensen, O., Donnelly, M. L., Francis, M. B., Sorg, J. A., and Shen, A.** 2015, PLoS Pathog., p. 11:e1005239.

26. *Global Analysis of the Sporulation Pathway of Clostridium difficile.* **Kelly A. Fimlaid, Jeffrey P. Bond, Kristin C. Schutz, Emily E. Putnam, Jacqueline M. Leung, Trevor D. Lawley, and Aimee Shen.** s.l. : PLoS Genet., 2013, Vol. e1003660.

27. *Sensitizing Clostridium difficile spores with germinants on skin and environmental surfaces represents a new strategy for reducing spores via ambient mechanisms.* **Nerandzic, M. M., and Donskey, C. J.** 2017, Pathog. Immun. , pp. 2, 404–421.

28. *Intestinal calcium and bile salts facilitate germination of Clostridium difficile spores.* **Kochan, T. J., Somers, M. J., Kaiser, A. M., Shoshiev, M. S., Hagan, A. K., Hastie, J. L., et al.** 2017, PLoS Pathog., p. 13:e1006443.

29. *Clostridioides difficile Biology: Sporulation, Germination, and Corresponding Therapies for C. difficile Infection.* **Sun, Duolong Zhu Joseph A. Sorg and Xingmin.** s.l. : Front Cell Infect Microbiol. , 2018.

30. *Interactions between Bacteria and Bile Salts in the Gastrointestinal and Hepatobiliary Tracts.* **Casadesús, Verónica Urdaneta and Josep.** s.l. : Front. Med., 2017 .

31. *A Gut Odyssey: The Impact of the Microbiota on Clostridium difficile Spore Formation and Germination.* **Shen, Aimee.** 11(10): e1005157. https://doi.org/10.1371/journal.ppat.1005157, s.l. : PLoS Pathog , 2015.

32. *Spore germination.* **Setlow, P.** 2003, Curr. Opin. Microbiol., pp. 6, 550–556.

33. *Recurrent Clostridium difficile infection: a review of risk factors, treatments, and outcomes.* **Johnson, S.** 2009, J. Infect., pp. 58, 403–410.

34. *Contamination of a hospital environment byClostridium difficile.* **Maury Ellis Mulligan, Rial D. Rolfe, Sydney M. Finegold & W. Lance George.** 1979, Current Microbiology, pp. 173-175.

35. *Recurrent Clostridioides difficile infection: Recognition, management, prevention.* **Constantine Tsigrelis, MD.** 2020, Cleveland Clinic Journal of Medicine, pp. 87 (6) 347-359;.

36. *Updates in Treatment of Recurrent Clostridium difficile Infection.* **Tanveer Singh, Prabhjot Bedi, Karandeep Bumrah, Jeevandeep Singh, Manoj Rai, and Susmitha Seelam.** 2019, J Clin Med Res, pp. 11(7): 465–471.

37. *Recurrent Clostridium difficile diarrhea: characteristics of and risk factors for patients enrolled in a prospective, randomized, double-blinded trial.* **Fekety R, McFarland LV, Surawicz CM, Greenberg RN, Elmer GW, Mulligan ME.** 1997, Clin Infect Dis., pp. 24(3):324-33.

38. *Metabolism of bile salts in mice influences spore germination in Clostridium difficile.* **Giel JL, Sorg JA, Sonenshein AL, Zhu J.** 2010, PLoS One., p. 5(1):e8740.

39. *The toxigenic element of Clostridium difficile strain VPI 10463.* **Hammond GA, Johnson JL.** 1995, Microb Pathog., pp. 19(4):203-13.

40. *Definition of the single integration site of the pathogenicity locus in Clostridium difficile.* **Braun V, Hundsberger T, Leukel P, Sauerborn M, von Eichel-Streiber C.** 1996, Gene, pp. 181(1-2):29-38.

41. *Production of a complete binary toxin (actin-specific ADP-ribosyltransferase) by Clostridium difficile CD196.* **Perelle S, Gibert M, Bourlioux P, Corthier G, Popoff MR.** 1997, Infect Immun., pp. 65(4):1402-7.

42. *Binary toxin production in Clostridium difficile is regulated by CdtR, a LytTR family response regulator.* **Carter GP, Lyras D, Allen DL, Mackin KE, Howarth PM, O'Connor JR, Rood JI.** 2007, J Bacteriol. , pp. 189(20):7290-301.

43. *Crystal structure and site-directed mutagenesis of enzymatic components from Clostridium perfringens iota-toxin.* **Tsuge H, Nagahama M, Nishimura H, Hisatsune J, Sakaguchi Y, Itogawa Y, Katunuma N, Sakurai J.** 2003, J Mol Biol., pp. 325(3):471-83.

44. *Biological activities of toxins A and B of Clostridium difficile.* **Lyerly DM, Lockwood DE, Richardson SH, Wilkins TD.** 1982, Infect Immun., pp. 35(3):1147-50.

45. *Integration of metabolism and virulence in Clostridium difficile.* **Bouillaut L, Dubois T, Sonenshein AL, Dupuy B.** 2015, Res Microbiol., pp. 166(4):375-83.

46. *Secretion of Clostridium difficile toxins A and B requires the holin-like protein TcdE.* **Govind R, Dupuy B.** 2012, PLoS Pathog., p. 8(6):e1002727.

47. *Toxin A-negative, toxin B-positive Clostridium difficile.* **Drudy D, Fanning S, Kyne L.** 2007, Int J Infect Dis., pp. 11(1):5-10.

48. *Clostridium difficile: New Insights into the Evolution of the Pathogenicity Locus.* **Monot M, Eckert C, Lemire A, Hamiot A, Dubois T, Tessier C, Dumoulard B, Hamel B, Petit A, Lalande V, Ma L, Bouchier C, Barbut F, Dupuy B.** 2015, Sci Rep., p. 5():15023.

49. *Transcription analysis of the genes tcdA-E of the pathogenicity locus of Clostridium difficile.* **Hundsberger T, Braun V, Weidmann M, Leukel P, Sauerborn M, von Eichel-Streiber C.** 1997, Eur J Biochem., pp. 244(3):735-42.

50. *Positive regulation of Clostridium difficile toxins.* **Moncrief JS, Barroso LA, Wilkins TD.** 1997, Infect Immun. , pp. 65(3):1105-8.

51. *Regulation of toxin synthesis in Clostridium difficile by an alternative RNA polymerase sigma factor.* **Mani N, Dupuy B.** 2001, Proc Natl Acad Sci U S A., pp. 98(10):5844-9.

52. *Evidence for holin function of tcdE gene in the pathogenicity of Clostridium difficile.* **Tan KS, Wee BY, Song KP.** 2001, J Med Microbiol., pp. 50(7):613-619.

53. *Glucosylation of Rho proteins by Clostridium difficile toxin B.* **Just I, Selzer J, Wilm M, von Eichel-Streiber C, Mann M, Aktories K.** 1995, Nature., pp. 375(6531):500-3.

54. *Comparative sequence analysis of the Clostridium difficile toxins A and B.* **von Eichel-Streiber C, Laufenberg-Feldmann R, Sartingen S, Schulze J, Sauerborn M.** 1992, Mol Gen Genet., pp. 233(1-2):260-8.

55. *Bacterial cytotoxins: targeting eukaryotic switches.* **Aktories K, Barbieri JT.** 2005, Nat Rev Microbiol, pp. 3(5):397-410.

56. *Auto-catalytic cleavage of Clostridium difficile toxins A and B depends on cysteine protease activity.* **Egerer M, Giesemann T, Jank T, Satchell KJ, Aktories K.** 2007, J Biol Chem., pp. 282(35):25314-21.

57. *Glucosylation Drives the Innate Inflammatory Response to Clostridium difficile Toxin A.* **Cowardin CA, Jackman BM, Noor Z, Burgess SL, Feig AL, Petri WA Jr.** 2016, Infect Immun., pp. 84(8):2317-2323.

58. *Clostridium difficile toxin A carries a C-terminal repetitive structure homologous to the carbohydrate binding region of streptococcal glycosyltransferases.* **von Eichel-Streiber C, Sauerborn M.** 1990, Gene, pp. 96(1):107-13.

59. *The combined repetitive oligopeptides of clostridium difficile toxin A counteract premature cleavage of the glucosyl-transferase domain by stabilizing protein conformation.* **Olling A, Hüls C, Goy S, Müller M, Krooss S, Rudolf I, Tatge H, Gerhard R.** 2014, Toxins (Basel), pp. 6(7):2162-76.

60. *Cellular uptake of Clostridium difficile toxin B. Translocation of the N-terminal catalytic domain into the cytosol of eukaryotic cells.* **Pfeifer G, Schirmer J, Leemhuis**

J, Busch C, Meyer DK, Aktories K, Barth H. 2003, J Biol Chem, pp. 278(45):44535-41.

61. *Autocatalytic cleavage of Clostridium difficile toxin B.* **Reineke J, Tenzer S, Rupnik M, Koschinski A, Hasselmayer O, Schrattenholz A, Schild H, von Eichel-Streiber C.** 2007, Nature, pp. 446(7134):415-9.

62. *The chaperonin TRiC/CCT is essential for the action of bacterial glycosylating protein toxins like Clostridium difficile toxins A and B.* **Marcus Steinemann, Andreas Schlosser, Thomas Jank, and Klaus Aktoriesa.** 2018, Proc Natl Acad Sci U S A., pp. 115(38): 9580–9585.

63. *Clostridium difficile toxin B activates dual caspase-dependent and caspase-independent apoptosis in intoxicated cells.* **Qa'Dan M, Ramsey M, Daniel J, Spyres LM, Safiejko-Mroczka B, Ortiz-Leduc W, Ballard JD.** 2002, Cell Microbiol., pp. 4(7):425-34.

64. *Rho protein inactivation induced apoptosis of cultured human endothelial cells.* **Hippenstiel S, Schmeck B, N'Guessan PD, Seybold J, Krüll M, Preissner K, Eichel-Streiber CV, Suttorp N.** 2002, Am J Physiol Lung Cell Mol Physiol., pp. 283(4):L830-8.

65. *R-ketorolac Targets Cdc42 and Rac1 and Alters Ovarian Cancer Cell Behaviors Critical for Invasion and Metastasis.* **Yuna Guo, Shelby Ray Kenney, Carolyn Y. Muller, Sarah Adams, Teresa Rutledge, Elsa Romero, Cristina Murray-Krezan, Rytis Prekeris, Larry A. Sklar, Laurie G. Hudson, and Angela Wandinger-Ness.** 2015, Mol Cancer Ther, pp. 14(10): 2215–2227.

66. *G Protein regulation of MAPK networks.* **Dhanasekaran, ZG Goldsmith and DN.** 2007, Oncogene, pp. 26, 3122–3142.

67. *Enteric glial cells counteract Clostridium difficile Toxin B through a NADPH oxidase/ROS/JNK/caspase-3 axis, without involving mitochondrial pathways.* **Macchioni L, Davidescu M, Fettucciari K, Petricciuolo M, Gatticchi L, Gioè D, Villanacci V, Bellini M, Marconi P, Roberti R, Bassotti G, Corazzi L.** 2017, Sci Rep., p. 7():45569.

68. *Functional defects in Clostridium difficile TcdB toxin uptake identify CSPG4 receptor-binding determinants.* **Pulkit Gupta, Zhifen Zhang, Seiji N Sugiman-Marangos, John Tam, Swetha Raman, Jean-Phillipe Julien, Heather K Kroh, D Borden Lacy, Nicholas Murgolo, Kavitha Bekkari, Alex G Therien, Lorraine D Hernandez, Roman A Melnyk.** 2017, J Biol Chem, pp. 292(42):17290-17301.

69. *Evidence for a modular structure of the homologous repetitive C-terminal carbohydrate-binding sites of Clostridium difficile toxins and Streptococcus mutans*

glucosyltransferases. **von Eichel-Streiber C, Sauerborn M, Kuramitsu HK.** 1992, J Bacteriol., pp. 174(20):6707-10.

70. *Crystal structure of receptor-binding C-terminal repeats from Clostridium difficile toxin A.* **Ho JG, Greco A, Rupnik M, Ng KK.** 2005, Proc Natl Acad Sci U S A., pp. 102(51):18373-8.

71. *Structural organization of the functional domains of Clostridium difficile toxins A and B.* **Pruitt RN, Chambers MG, Ng KK, Ohi MD, Lacy DB.** 2010, Proc Natl Acad Sci U S A., pp. 107(30):13467-72.

72. *Cell surface binding site for Clostridium difficile enterotoxin: evidence for a glycoconjugate containing the sequence Gal alpha 1-3Gal beta 1-4GlcNAc.* **Krivan HC, Clark GF, Smith DF, Wilkins TD.** 1986, Infect Immun., pp. 53(3):573-81.

73. *Carbohydrate recognition by Clostridium difficile toxin A.* **Greco A, Ho JG, Lin SJ, Palcic MM, Rupnik M, Ng KK.** 2006, Nat Struct Mol Biol., pp. 13(5):460-1.

74. *Rabbit sucrase-isomaltase contains a functional intestinal receptor for Clostridium difficile toxin A.* **Pothoulakis C, Gilbert RJ, Cladaras C, Castagliuolo I, Semenza G, Hitti Y, Montcrief JS, Linevsky J, Kelly CP, Nikulasson S, Desai HP, Wilkins TD, LaMont JT.** 1996, J Clin Invest., pp. 98(3):641-9.

75. *gp96 Is a Human Colonocyte Plasma Membrane Binding Protein for Clostridium difficile Toxin A.* **Xi Na, Ho Kim, Mary P. Moyer, Charalabos Pothoulakis, J. Thomas LaMont.** 2008, Infect Immun , pp. 76:2862–2871.

76. *A family of clostridial and streptococcal ligand-binding proteins with conserved C-terminal repeat sequences.* **Wren, B. W.** 1991, Mol. Microbiol, pp. 5:797-803.

77. *Chondroitin sulfate proteoglycan 4 functions as the cellular receptor for Clostridium difficile toxin B.* **Pengfei Yuan, Hongmin Zhang, Changzu Cai, Shiyou Zhu, Yuexin Zhou, Xiaozhou Yang, Ruina He, Chan Li, Shengjie Guo, Shan Li, Tuxiong Huang, Gregorio Perez-Cordon, Hanping Feng, Wensheng Wei.** 2015, Cell Res, pp. 25(2):157-68.

78. *Identification of an epithelial cell receptor responsible for Clostridium difficile TcdB-induced cytotoxicity.* **Michelle E LaFrance, Melissa A Farrow, Ramyavardhanee Chandrasekaran, Jinsong Sheng, Donald H Rubin, D Borden Lacy.** 2015, Proc Natl Acad Sci U S A, pp. 112(22):7073-8.

79. *Frizzled proteins are colonic epithelial receptors for C. difficile toxin B.* **Liang Tao, Jie Zhang, Paul Meraner, Alessio Tovaglieri, Xiaoqian Wu, Ralf Gerhard, Xinjun Zhang, William B. Stallcup, Ji Miao, Xi He, Julian G. Hurdle, David T. Breault, Abraham L. Brass & Min Dong.** 2016, Nature, pp. 350–355.

80. *Structural basis for recognition of frizzled proteins by Clostridium difficile toxin B.* **Peng Chen, Liang Tao, Tianyu Wang, Jie Zhang, Aina He, Kwok-Ho Lam, Zheng Liu, Xi He, Kay Perry, Min Dong, Rongsheng Jin.** 2018, Science , pp. 360(6389):664-669.

81. *Receptor Binding Domains of TcdB from Clostridioides difficile for Chondroitin Sulfate Proteoglycan-4 and Frizzled Proteins Are Functionally Independent and Additive.* **Daniel Henkel, Helma Tatge, Dennis Schöttelndreier, Liang Tao, Min Dong and Ralf Gerhard.** https://doi.org/10.3390/toxins12120736, s.l. : Toxins , 2020.

82. *A Bacterial Effector Reveals the V-ATPase-ATG16L1 Axis that Initiates Xenophagy.* **Yue Xu, Ping Zhou, Sen Cheng, Qiuhe Lu, Kathrin Nowak, Ann-Katrin Hopp, Lin Li, Xuyan Shi, Zhiwei Zhou, Wenqing Gao, Da Li, Huabin He, Xiaoyun Liu, Jingjin Ding, Michael O. Hottiger, and Feng Shao.** 2019, Cell, pp. 178, 552–566.

83. *The Ins and Outs of Anthrax Toxin.* **Sarah Friebe, F. Gisou van der Goot, and Jérôme Bürgi.** s.l. : Toxins (Basel), 2016 , Vol. doi: 10.3390/toxins8030069.

84. *Understanding the mode of action of diphtheria toxin: a perspective on progress during the 20th century.* **R.JCollier.** 11, s.l. : Toxicon , 2001, Vol. 39.

85. *Variations in TcdB Activity and the Hypervirulence of Emerging Strains of Clostridium difficile.* **Jordi M. Lanis, Soumitra Barua, and Jimmy D. Ballard.** 2010, PLoS Pathog, p. 6(8): e1001061.

86. *Translocation domain mutations affecting cellular toxicity identify the Clostridium difficile toxin B pore.* **Zhifen Zhang, Minyoung Park, John Tam, Anick Auger, Greg L. Beilhartz, D. Borden Lacy, and Roman A. Melnyka.** 2014, Proc Natl Acad Sci U S A. , pp. 111(10): 3721–3726.

87. *Cholesterol-dependent pore formation of Clostridium difficile toxin A.* **Torsten Giesemann, Thomas Jank, Ralf Gerhard, Elke Maier, Ingo Just, Roland Benz, Klaus Aktories.** 2006, J Biol Chem, pp. 281(16):10808-15.

88. *Crystal structure of Clostridium difficile toxin A.* **Nicole M Chumbler, Stacey A Rutherford, Zhifen Zhang, Melissa A Farrow, John P Lisher, Erik Farquhar, David P Giedroc, Benjamin W Spiller, Roman A Melnyk, D Borden Lacy.** 2016, Nat Microbiol , p. 1:15002.

89. *Structure of the full-length Clostridium difficile toxin B.* **Peng Chen, Kwok-Ho Lam, Zheng Liu, Frank A Mindlin, Baohua Chen, Craig B Gutierrez, Lan Huang, Yongrong Zhang, Therwa Hamza, Hanping Feng, Tsutomu Matsui, Mark E Bowen, Kay Perry, Rongsheng Jin.** 2019, Nat Struct Mol Biol , pp. 26(8):712-719.

90. *Structural determinants for membrane insertion, pore formation and translocation of Clostridium difficile toxin B.* **Aktories, Selda Genisyuerek Panagiotis Papatheodorou Gregor Guttenberg Rolf Schubert Roland Benz Klaus.** 2011, molecular microbiology, pp. 1643-1654.

91. —.**Genisyuerek S., Papatheodorou P., Guttenberg G., Schubert R., Benz R., Aktories K.** 2011, Mol. Microbiol, pp. 79, 1643–1654.

92. *Clostridium difficile glucosyltransferase toxin B-essential amino acids for substrate binding.* **Jank T, Giesemann T, Aktories K J.** 2007, Jank T, Giesemann T, Aktories K J Biol Chem. , pp. 282(48):35222-31.

93. *Back in the water: the return of the inositol phosphates.* **Schell, Robin F. Irvine & Michael J.** 2001, Nature Reviews Molecular Cell Biology, pp. 327–338.

94. *Inositol Hexakisphosphate-Induced Autoprocessing of Large Bacterial Protein Toxins.* **Martina Egerer, Karla J. F. Satchell.** https://doi.org/10.1371/journal.ppat.1000942, s.l. : PLoS Pathog , 2010.

95. *Structure-function analysis of inositol hexakisphosphate-induced autoprocessing in Clostridium difficile toxin A.* . **Pruitt, R. N., Chagot, B., Cover, M., Chazin, W. J., Spiller, B., and Lacy, D. B.** 2009, J. Biol. Chem. , pp. 284, 21934–21940.

96. *Defining an allosteric circuit in the cysteine protease domain of Clostridium difficile toxins.* **Shen A, Lupardus PJ, Gersch MM, Puri AW, Albrow VE, Garcia KC, Bogyo M.** 2011, Nat Struct Mol Biol., pp. 18(3):364-71.

97. *The combined repetitive oligopeptides of Clostridium difficile toxin A counteract premature cleavage of the glucosyl-transferase domain by stabilizing protein conformation.* . **Olling A, Huls C, Goy Set al.** 2014, Toxins, pp. 6:2162–76.

98. *Comparative genome and phenotypic analysis of Clostridium difficile 027 strains provides insight into the evolution of a hypervirulent bacterium.* . **Stabler, R. A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., Lawley, T. D., Sebaihia, M., Quail, M. A., Rose, G., Gerding, D. N., Gibert, M., Popoff, M. R., Parkhill, J., Dougan, G., and Wren, B. W.** 2009, Genome Biol. , pp. 10, R102.

99. *Small GTPases: Structure, biological function and its interaction with nanoparticles.* **Siyang Song, Wenshu Cong, Shurong Zhou, Yujie Shi, Wenbing Dai, Hua Zhang, Xueqing Wang, Bing He, Qiang Zhang.** 2019, Asian Journal of Pharmaceutical Sciences, pp. 30-39.

100. *Impact of amino acids 22-27 of Rho-subfamily GTPases on glucosylation by the large clostridial cytotoxins TcsL-1522, TcdB-1470 and TcdB-8864.* **Müller S, von Eichel-Streiber C, Moos M.** 1999, Eur J Biochem. , pp. 266(3):1073-80.

101. *Structural basis for the function of Clostridium difficile toxin B.* **Reinert D. J., Jank T., Aktories K., Schulz G. E.** 2005, J. Mol. Biol., pp. 351, 973–981.

102. *Primary symptomless colonisation by Clostridium difficile and decreased risk of subsequent diarrhoea.* **Shim, J. K., S. Johnson, M. H. Samore, D. Z. Bliss, and D. N. Gerding.** 1998, Lancet, pp. 351:633-636.

103. *Identification of toxin A-negative, toxin B-positive Clostridium difficile by PCR.* **Kato, H., N. Kato, K. Watanabe, N. Iwai, H. Nakamura, T. Yamamoto, K. Suzuki, S. M. Kim, Y. Chong, and E. B. Wasito.** 1998, J. Clin. Microbiol, pp. 36:2178-2182.

104. *Fatal pseudomembranous colitis associated with a variant Clostridium difficile strain not detected by toxin A immunoassay.* **Johnson, S., S. A. Kent, K. J. O'Leary, M. M. Merrigan, S. P. Sambol, L. R. Peterson, and D. N. Gerding.** 2001, Ann. Intern. Med, pp. 135:434-438.

105. *Clostridium difficile--more difficult than ever.* **Kelly CP, LaMont JT.** 2008, N Engl J Med. , pp. 359(18):1932-40.

106. *Treatment of recurrent Clostridium difficile colitis: A narrative review.* **R, Hopkins R and Wilson.** 2018, Gastroenterol Rep (Oxf), pp. 6(1):21-8.

107. *Comparison of vancomycin, teicoplanin, metronidazole, and fusidic acid for the treatment of Clostridium difficile-associated diarrhea.* **Wenisch C, Parschalk B, Hasenhündl M, Hirschl AM, Graninger W.** 1996, Clin Infect Dis. , pp. 22(5):813-8.

108. *Host S-nitrosylation inhibits clostridial small molecule-activated glucosylating toxins.* **Tor C Savidge, Petri Urvil, Numan Oezguen, Kausar Ali, Aproteem Choudhury, Vinay Acharya, Irina Pinchuk, Alfredo G Torres, Robert D English, John E Wiktorowicz, Michael Loeffelholz, Raj Kumar, Lianfa Shi, Weijia Nie, Werner Braun, Bo Herman, Alfred Hausladen, Hanping Feng, Jonathan S Stamler, and Charalabos Pothoulakis.** 2011, Nat Med. , pp. 17(9): 1136–1141.

109. *Guidelines for diagnosis, treatment, and prevention of Clostridium difficile infections.* **Surawicz CM, Brandt LJ, Binion DG, Ananthakrishnan AN, Curry SR, Gilligan PH, McFarland LV, Mellow M, Zuckerbraun BS.** 2013, Am J Gastroenterol, pp. 108(4):478-98.

110. *A randomized, double-blind, placebo-controlled pilot study to assess the ability of rifaximin to prevent recurrent diarrhoea in patients with Clostridium difficile infection.* **Garey KW, Ghantoji SS, Shah DN, Habib M, Arora V, Jiang ZD, et al.** 2011, Journal of Antimicrobial Chemotherapy, pp. 66(12):2850-5.

111. *Bezlotoxumab for Prevention of Recurrent Clostridium difficile Infection.* **Wilcox MH, Gerding DN, Poxton IR, Kelly C, Nathan R, Birch T, Cornely OA, Rahav G,**

Bouza E, Lee C, Jenkin G, Jensen W, Kim YS, Yoshida J, Gabryelski L, Pedley A, Eves K, Tipping R, Guris D, Kartsonis N, Dorr MB, MODIFY I and MODIFY II Investigators. 2017, N Engl J Med., pp. 376(4):305-317.

112. *Expression of recombinant Clostridium difficile toxin A and B in Bacillus megaterium.* **Guilin Yang, Boping Zhou, Jufang Wang, Xiangyun He, Xingmin Sun, Weijia Nie, Saul Tzipori, and Hanping Feng.** s.l. : BMC Microbiol. , 2008.

113. *Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM.* **Xueming Li, Paul Mooney, Shawn Zheng, Christopher R Booth, Michael B Braunfeld, Sander Gubbens, David A Agard & Yifan Cheng.** s.l. : Nature Methods , 2013, Vols. 10, pages584–590(2013).

114. *Gctf: Real-time CTF determination and correction.* **Zhang, Kai.** s.l. : J Struct Biol, 2016, Vols. 193(1): 1–12.

115. *Selection and characterization of ultrahigh potency designed ankyrin repeat protein inhibitors of C. difficile toxin B.* **Rudo Simeon, Mengqiu Jiang, Ana M Chamoun-Emanuelli, Hua Yu, Yongrong Zhang, Ran Meng, Zeyu Peng, Joanita Jakana, Junjie Zhang, Hanping Feng, Zhilei Chen.** s.l. : PLoS Biol, 2019 , Vol. 17(6):e3000311.

116. *RELION: Implementation of a Bayesian approach to cryo-EM structure determination.* **H.W.Scheres, Sjors.** 3, s.l. : Journal of Structural Biology, 2012, Vol. 180.

117. *Quantifying the Local Resolution of Cryo-EM Density Maps.* **A. Kucukelbir, F.J. Sigworth, and H.D. Tagare.** 1, s.l. : Nature Methods. , 2014, Vol. 11.

118. *UCSF Chimera - A Visualization System for Exploratory Research and Analysis.* **Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E.** s.l. : J. Comput. Chem., 2004, Vols. 25(13):1605-1612.

119. *Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics.* **Leonardo G. Trabuco, Elizabeth Villa, Kakoli Mitra, Joachim Frank, and Klaus Schulten.** s.l. : Structure, 2008, Vols. 16, 673-683.

120. *Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix* . **D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L.-W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D.** s.l. : Acta Cryst, 2019, Vols. D75, 861-877.

121. *Coot: model-building tools for molecular graphics.* **Emsley P, Cowtan K.** s.l. : Acta Crystallogr, (2004), Vols. D60, 2126-2132.

122. *Protein structure prediction using Rosetta.* **Rohl CA, Strauss CE, Misura KM, Baker D.** s.l. : Methods in Enzymology, 2003, Vols. 383:66-93.

123. *I-TASSER: a unified platform for automated protein structure and function prediction.* **Ambrish Roy, Alper Kucukural & Yang Zhang.** s.l. : Nature Protocols , 2010, Vols. 5, pages725–738.

124. *Protein structure prediction and analysis using the Robetta server.* **David E. Kim, Dylan Chivian, and David Baker.** 32(Web Server issue): W526–W531., s.l. : Nucleic Acids Res., 2004 .

125. *Chondroitin Sulfate Proteoglycan 4 and Its Potential As an Antibody Immunotherapy Target across Different Tumor Types.* **Kristina M. Ilieva, Anthony Cheung, Silvia Mele, Giulia Chiaruttini, Silvia Crescioli, Merope Griffin, Mano Nakamura, James F. Spicer, Sophia Tsoka, Katie E. Lacy, Andrew N. J. Tutt, and Sophia N. Karagiannis.** s.l. : Front Immunol, 2017.

126. **Joseph W. Alvin, D. Borden Lacy.** Structure Function Studies of Large Clostridial Cytotoxins. [book auth.] Stiles B., Alape-Girón A., Dubreuil J., Mandal M. Gopalakrishnakone P. *Microbial Toxins.* . s.l. : Toxinology. Springer, Dordrecht., 2017.

127. *Structural determinants for membrane insertion, pore formation and translocation of Clostridium difficile toxin B.* **Aktories, Selda Genisyuerek Panagiotis Papatheodorou Gregor Guttenberg Rolf Schubert Roland Benz Klaus.** s.l. : molelcular microbiology, 2011.

128. *Low pH-induced formation of ion channels by Clostridium difficile toxin B in target cells.* . **H Barth, G Pfeifer, F Hofmann, E Maier, R Benz, K Aktories.** s.l. : Journal of biological chemistry, 2001.

129. *cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination.* **Ali Punjani, John L Rubinstein, David J Fleet & Marcus A Brubaker.** s.l. : nature methods, 2017, Vols. 14, pages290–296.

130. *Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix* . **D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L.-W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D.** s.l. : Acta Cryst, (2019), Vols. D75, 861-877.

131. *Efficient cell delivery mediated by lipid-specific endosomal escape of supercharged branched peptides.* **Brock, D.J., Kustigian, L., Jiang, M., Graham, K., Wang, T.Y., Erazo-Oliveras, A., Najjar, K., Zhang, J., Rye, H., and Pellois, J.P.** s.l. : Traffic, 2018, Vols. 19, 421-435.

132. *Burst analysis spectroscopy: a versatile single-particle approach for studying distributions of protein aggregates and fluorescent assemblies.* **Puchalla, J., Krantz, K., Austin, R., and Rye, H.** s.l. : Proceedings of the National Academy of Sciences of the United States of America, 2008, Vols. 105, 14400-14405.

133. *Single particle fluorescence burst analysis of epsin induced membrane fission.* **Brooks, A., Shoup, D., Kustigian, L., Puchalla, J., Carr, C.M., and Rye, H.S.** s.l. : PloS one , 2015, Vols. 10, e0119563.

134. *Burst analysis spectroscopy: A versatile single-particle approach for studying distributions of protein aggregates and fluorescent assemblies.* **Jason Puchalla, Kelly Krantz, Robert Austin, and Hays Rye.** s.l. : PNAS , 2008 , Vols. 105 (38) 14400-14405.

135. *Cryo-EM analysis of a membrane protein embedded in the liposome* . **Xia Yao, Xiao Fan, and Nieng Yan.** s.l. : PNAS, 2020, Vols. 117 (31) 18497-18503.

136. *Current status of Clostridium difficile infection epidemiology.* . **Lessa FC, Gould CV, McDonald LC.** s.l. : Clin Infect Dis., 2012, Vols. 55 Suppl 2():S65-70.

137. *Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins.* . **Binz HK, Stumpp MT, Forrer P, Amstutz P, Plückthun A.** s.l. : J Mol Biol., 2003 , Vols. 332(2):489-503.

138. *Designed ankyrin repeat proteins: a novel tool for testing epidermal growth factor receptor 2 expression in breast cancer.* . **Theurillat JP, Dreier B, Nagy-Davidescu G, Seifert B, Behnke S, Zürrer-Härdi U, Ingold F, Plückthun A, Moch H.** s.l. : Mod Pathol., 2010, Vols. 23(9):1289-97.

139. *Efficient selection of DARPins with sub-nanomolar affinities using SRP phage display.* **Steiner D, Forrer P, Plückthun A.** s.l. : J Mol Biol. , 2008 , Vols. 382(5):1211-27.

140. *Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6.* . **Grant T, Grigorieff N.** s.l. : Elife. , 2015 , Vol. 4():e06980.

141. *CTFFIND4: Fast and accurate defocus estimation from electron micrographs.* . **Rohou A, Grigorieff N.** s.l. : J Struct Biol. , 2015 , Vols. 192(2):216-21.

142. *EMAN: semiautomated software for high-resolution single-particle reconstructions.* . **Ludtke SJ, Baldwin PR, Chiu W.** s.l. : J Struct Biol., 1999, Vols. 128(1):82-97.

143. *Bsoft: image processing and molecular modeling for electron microscopy.* **Heymann JB, Belnap DM.** s.l. : J Struct Biol., 2007, Vols. 157(1):3-18.

144. *Bispecific designed ankyrin repeat proteins (DARPins) targeting epidermal growth factor receptor inhibit A431 cell proliferation and receptor recycling.* **Boersma YL, Chao G, Steiner D, Wittrup KD, Plückthun A.** s.l. : J Biol Chem. , 2011 , Vols. 286(48):41273-85.

145. *Lectin activity of the TcdA and TcdB toxins of Clostridium difficile.* **Lauren E. Hartley-Tassell, Milena M. Awad , Kate L. Seib , Maria Scarselli, Silvana Savino, Joe Tiralongo , Dena Lyras , Christopher J. Day , Michael P. Jennings.** s.l. : Infect. Immun, 2018, Vols. doi:10.1128/IAI.00676-18.

146. *Clostridium difficile Toxin A Undergoes Clathrin-Independent, PACSIN2-Dependent Endocytosis.* **Chandrasekaran R, Kenworthy AK, Lacy DB.** s.l. : PLoS Pathog, 2016, Vol. 12(12): e1006070. https://doi.org/10.1371/journal.ppat.1006070.

147. *Clostridium difficile toxins A and B: Receptors, pores, and translocation into cells.* **Kathleen E Orrell, Zhifen Zhang, Seiji N Sugiman-Marangos, Roman A Melnyk.** s.l. : Crit Rev Biochem Mol Biol , 2017, Vols. 52(4):461-473.

148. *Endocytosis via caveolae: alternative pathway with distinct cellular compartments to avoid lysosomal degradation?* **Botos, Anna L Kiss* and Erzsébet.** s.l. : J Cell Mol Med., 2009 , Vols. 13(7): 1228–1237.

149. *Functional properties of the carboxy-terminal host cell-binding domains of the two toxins, TcdA and TcdB, expressed by Clostridium difficile.* **Tanis Dingle, Stefanie Wee, George L Mulvey, Antonio Greco, Elena N Kitova, Jiangxiao Sun, Shuangjun Lin, John S Klassen, Monica M Palcic, Kenneth K S Ng, Glen D Armstrong.** 9, s.l. : Glycobiology, 2008, Vol. 18.

APPENDIX A

*GROEL ACTIVELY STIMULATESF FOLDING OF THE ENDOGENOUS SUBSTRATE PROTEIN PEPQ

ARTICLE

Received 2 Oct 2016 | Accepted 13 May 2017 | Published 30 Jun 2017

DOI: 10.1038/ncomms15934 OPEN

GroEL actively stimulates folding of the endogenous substrate protein PepQ

Jeremy Weaver[1,*,†], Mengqiu Jiang[1,2,*], Andrew Roth[1], Jason Puchalla[3], Junjie Zhang[1] & Hays S. Rye[1]

Many essential proteins cannot fold without help from chaperonins, like the GroELS system of *Escherichia coli*. How chaperonins accelerate protein folding remains controversial. Here we test key predictions of both passive and active models of GroELS-stimulated folding, using the endogenous *E. coli* metalloprotease PepQ. While GroELS increases the folding rate of PepQ by over 15-fold, we demonstrate that slow spontaneous folding of PepQ is not caused by aggregation. Fluorescence measurements suggest that, when folding inside the GroEL-GroES cavity, PepQ populates conformations not observed during spontaneous folding in free solution. Using cryo-electron microscopy, we show that the GroEL C-termini make physical contact with the PepQ folding intermediate and help retain it deep within the GroEL cavity, resulting in reduced compactness of the PepQ monomer. Our findings strongly support an active model of chaperonin-mediated protein folding, where partial unfolding of misfolded intermediates plays a key role.

[1] Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas 77845, USA. [2] State Key Laboratory of Biocontrol, School of Life Science, Sun Yat-sen University, Guangzhou, Guangdong 510275, China. [3] Department of Physics, Princeton University, Princeton, New Jersey 08544, USA. * These authors contributed equally to this work. † Present address: Division of Molecular and Cellular Biology, NICHD, National Institutes of Health, Bethesda, Maryland 20892, USA. Correspondence and requests for materials should be addressed to J.Z. (email: junjiez@tamu.edu) or to H.S.R. (email: haysrye@tamu.edu).

Folding is a highly error prone process for many large and essential cellular proteins. Misfolding and aggregation often overwhelm the delicate thermodynamic balance that drives a protein toward its native state. Throughout evolutionary history, living systems have solved this problem with specialized, ATP-powered machines known as molecular chaperones[1]. The Hsp60s or chaperonins are a central and essential family of the molecular chaperones, and the GroELS chaperonin system of *Escherichia coli* is one of the best studied examples[2–5]. GroEL is a homo-oligomer of 14, 57 kDa subunits, that is arranged in two, seven membered rings stacked back-to-back. Each ring contains a large, open, solvent-filled cavity[6]. The inner cavity surface of the uppermost domain (the apical domain) is lined with hydrophobic amino acids that capture non-native substrate proteins[7,8]. Substrate proteins that strictly depend upon GroEL for folding (so-called stringent substrate proteins) must be briefly enclosed within a complex formed by a GroEL ring and the smaller, ring-shaped co-chaperonin GroES[9–12]. Formation of the GroEL-GroES complex first requires that a GroEL ring bind ATP, which triggers a series of conformational rearrangements of the GroEL ring, permitting GroES to bind and resulting in the encapsulation of the substrate protein. Enclosure of the substrate protein beneath GroES results in ejection and confinement of the protein inside the enlarged GroEL-GroES chamber (a *cis* complex) and initiation of protein folding. Folding continues within the isolated GroEL-GroES cavity for a brief period, until the complex is disassembled and the substrate protein, folded or not, is released back into free solution[9–11,13,14].

Despite this detailed structural and functional knowledge, current models of GroEL-assisted folding remain divided into two general types based upon whether GroEL is presumed to act passively or actively[2,3,5,15]. Passive models, like the Anfinsen cage or infinite dilution model, postulate that protein folding is only enhanced by GroELS because folding intermediates are prevented from aggregating by isolating them within the protective environment of the GroELS chamber[2,15]. Purely passive models implicitly assume that the folding of GroEL-dependent proteins are constrained only by the aggregation propensity of on-pathway folding intermediates. Active GroEL folding models, by contrast, assume that stringent GroEL-substrate proteins can and do populate off-pathway, kinetically trapped states. In this view, GroELS stimulates protein folding because these kinetically trapped intermediates benefit not only from protection against aggregation but also from additional, and essential, corrective actions provided by the chaperonin[3,16]. The mechanism of this corrective action remains controversial, but has been suggested to come from either (1) repetitive unfolding and iterative annealing[17,18] or (2) smoothing of a substrate protein's free energy landscape as a result of confinement inside the GroEL-GroES cavity, where either steric constraints and/or interactions within the chamber prevent unproductive folding pathways in favour of productive ones[3,15,16].

Several stringent substrate proteins have been shown to display folding behaviour that is consistent with one or more predictions of active GroEL folding models[19–21]. Some of the most detailed analysis to date has been conducted with ribulose-1, 5-bisphosphate carboxylase oxygenase (RuBisCO) from *R. rubrum* and a double mutant of *E. coli* maltose binding protein (MBP)[19–24]. While highly suggestive, these studies nonetheless leave the importance of active folding unclear. General conclusions about the impact of active folding cannot be robustly drawn from such a small number of examples. In addition, in the case of RuBisCO, the mismatch between the biological source of the substrate protein (*Rhodospirillum rubrum*) and the chaperonin (*E. coli*), leaves the biological consequences of these findings open to interpretation. Similarly, in the case MBP, it was necessary to employ an engineered double-mutant of this protein in order to study GroEL-stimulated folding, because wild-type MBP neither interacts with, nor needs the chaperonin for folding in its natural biological context. Thus, a convincing demonstration of active folding assistance by GroEL of a stringent, endogenous *E. coli* substrate protein has remained elusive. A recent study on the assisted folding of the *E. coli* HTP synthase/lyase DapA sought to address this problem[25]. The results of this work suggested that DapA requires an active GroEL folding mechanism. However, a more recent study of DapA folding called key elements of this work into question[26].

In order to test the central predictions of passive and active models of chaperonin-mediated folding, we have re-examined the mechanism of GroELS-assisted protein folding using the biologically relevant, endogenous *E. coli* prolidase enzyme, PepQ. PepQ catalyses the hydrolysis of dipeptides that contain C-terminal proline residues[27,28]. It forms a homodimer, with each monomer (~50 kDa) built from two domains: a small, mixed α/β N-terminal domain and a pita-bread fold[29,30] C-terminal domain that contains the active site (Fig. 1a; ref. 27). Two independent proteomics studies predicted that PepQ requires the assistance of GroEL-GroES for folding *in vivo*[31,32]. In addition, PepQ is a member of a protein structural family that is not represented among the well-characterized GroEL-substrate proteins. Here, using a combination of enzymatic assays, single-molecule fluorescence techniques, and cryo-electron microscopy (EM), we demonstrate that GroEL actively alters the folding of PepQ. Initial capture of a kinetically trapped PepQ monomer by a GroEL ring results in substantial unfolding, a process that relies in part on a direct, physical interaction between the PepQ folding intermediate and the unstructured GroEL C-terminal tails. Subsequent encapsulation of the partially unfolded folding PepQ monomer within the GroEL/ES chamber fundamentally alters the folding trajectory of the protein, resulting in a faster and more efficient search for the native state.

Results

Slow spontaneous PepQ folding is not caused by aggregation. Upon dilution from chemical denaturant, PepQ folds spontaneously at room temperature (23 °C) to a final yield of 50–60% with an observed half-time of ~20 min (Fig. 1b). However, in the presence of the cycling GroEL-GroES system, PepQ folds with an observed half-time of ~1 min to a final yield of 80–90% (Fig. 1b). Encapsulation of PepQ within a non-cycling chaperonin complex, composed of the GroEL single-ring mutant SR1 and GroES, also results in accelerated refolding, consistent with previous observations from other GroEL-substrate proteins (Fig. 1b; refs 9,25,33,34). Inside the static SR1-ES cavity, PepQ folds at a rate similar to that observed with cycling wild type GroEL at 23 °C, although it displays a consistently lower yield. Thus, while PepQ does not require GroEL to fold, the chaperonin accelerates the folding rate of the enzyme by 15–20-fold, while increasing the native state yield by ~40%.

The observation of slow spontaneous folding, in combination with a decreased native state yield, suggests that PepQ folding is inhibited by non-productive side reactions like misfolding or aggregation. We therefore examined the fate of the PepQ that fails to reach the native state. While this material remains fully in solution, over time it loses the ability to fold productively, even with assistance from GroEL (Supplementary Fig. 1). Approximately half of the PepQ population becomes refractory to GroEL-mediated folding with a time constant that is similar to that observed for productive spontaneous folding. Because the non-native states of many chaperonin-dependent proteins are highly prone to aggregation, we sought to determine whether

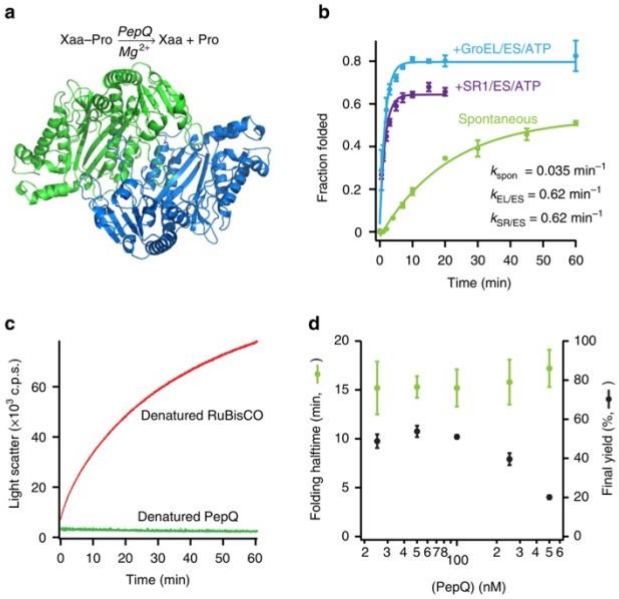


Figure 1 | Stimulated folding of PepQ by GroEL does not depend on large-scale suppression of aggregation. (**a**) The *E. coli* metalloprotease PepQ catalyses the hydrolysis of dipeptides containing C-terminal proline residues. The structure of the native PepQ homodimer (PDB ID: 4QR8) is shown, illustrating the pita-bread fold common to this enzyme family. (**b**) Refolding of PepQ was monitored by the recovery of enzymatic activity. PepQ was denatured in acid-urea and then diluted into either buffer alone (100 nM; spontaneous, green) or buffer containing GroEL (200 nM). The GroEL-PepQ binary complex was then supplemented with GroES (400 nM) and ATP (2 mM) to initiate folding (+GroEL/ES/ATP, blue). In a parallel experiment, denatured PepQ was bound to the single-ring mutant of GroEL, SR1 (300 nM), and refolded in the presence of GroES (600 nM) and ATP (2 mM; +SR1/ES/ATP, purple). Data were fit to a single-exponential rate law (solid lines), resulting in observed folding rate constants of $0.62 \pm 0.05\ min^{-1}$ for GroEL, $0.62 \pm 0.09\ min^{-1}$ for SR1 and $0.035 \pm 0.005\ min^{-1}$ for the spontaneous reaction. Error bars show the standard deviation of three independent experiments. (**c**) Large-scale aggregation of PepQ and RuBisCO was examined by static light scattering at 340 nm. PepQ (green) and RuBisCO (red) were each denatured in acid-urea and then separately diluted into buffer at 23 °C (100 nM final monomer). Each trace is the average of three separate experiments. (**d**) The rate and yield of spontaneous PepQ folding as a function of enzyme concentration is shown. Chemically denatured, wild type PepQ was diluted 50-fold into buffer to yield spontaneous folding reactions at the indicated final monomer concentration. The folding rate at each protein concentration (green) and native state yield (black) are shown. Error bars show the s.d. of three independent folding experiments.

inefficient PepQ folding was due to aggregation. We first examined the static light scattering of a spontaneous folding reaction in which PepQ was rapidly diluted from denaturant into refolding buffer. Surprisingly, PepQ displayed no significant increase in light scattering, even after 1 h of incubation at 23 °C (Fig. 1c). By contrast, *R. rubrum* RuBisCO, a stringent GroEL-substrate protein well known to aggregate at 23 °C (refs 19,35,36) showed a rapid and substantial increase in light scattering under the same conditions (Fig. 1c). These observations indicate that denatured PepQ does not form high concentrations of large aggregates, at least under the conditions of the spontaneous folding assay. However, PepQ could form inhibitory aggregates that are too small or rare to be well detected by light scattering. If true, the observed rate of spontaneous PepQ folding should be a sensitive function of the total protein concentration. Strikingly, over a concentration range from 25 to 500 nM, the observed half-time of PepQ folding remained unchanged, although we did observe a decrease in the native state yield as the protein concentration was increased above 250 nM (Fig. 1d).

The concentration independence of the PepQ folding rate suggested that the slowness of spontaneous folding is not caused by inhibitory aggregation. To further test this conclusion, we examined PepQ folding at low protein concentrations using a set of fluorescence-based assays. We first introduced a surface-exposed Cys residue into the first helix of the PepQ N-terminal domain (A24C), which permitted unique attachment of exogenous fluorescent probes (Supplementary Fig. 2A). Importantly, PepQ labelled at position 24 with small dyes like IAEDANS (PepQ-24ED), fluorescein (PepQ-24F), Oregon Green (PepQ-24OG) or tetramethyl rhodamine (PepQ-24TMR) displayed no apparent alteration in enzymatic activity or stability. Spontaneous folding of the PepQ-24ED variant displayed no significant difference compared to wild type PepQ, and the PepQ-24F, PepQ-24OG, PepQ-24TMR variants folded only slightly more slowly (~30%; Supplementary Fig. 2B). We used these labelled PepQ variants in an intermolecular Förster Resonance Energy Transfer (FRET) assay designed to examine aggregate formation during spontaneous PepQ folding[27]. In this assay, two differently labelled PepQ monomers were employed: PepQ-24ED as the donor and PepQ-24F as the acceptor. In the native PepQ dimer, these sites are positioned too far apart for Förster coupling (Supplementary Fig. 2A), so that any observed

FRET signal should report primarily on aggregate formation. When the two PepQ samples were mixed, denatured and diluted together into refolding buffer at 50 °C, formation of PepQ aggregates was readily observed as a robust FRET signal (80% FRET efficiency; Supplementary Fig. 2C). Surprisingly, when the same experiment was conducted under spontaneous folding conditions at 23 °C, the observed FRET efficiency was less than 4%, suggesting a lack of significant aggregation (Supplementary Fig. 2C).

We next examined PepQ folding and aggregation at extremely low protein concentrations using single-molecule detection techniques. First, samples of PepQ-24TMR were denatured in acid-urea and spontaneous folding was initiated by rapid dilution (50-fold) into refolding buffer at 23 °C, yielding a final monomer concentration of 2 nM. This sample was allowed to fold spontaneously at 23 °C and samples were removed and mixed with a large excess of GroEL at different time points. Excess GroEL was added to both quench the folding reaction and increase the effective diffusion time of uncommitted PepQ monomers, which were bound by the much larger GroEL tetradecamer[22,25]. Fluorescence correlation spectra (FCS) were then acquired for each time point and the fraction of folded versus non-native PepQ was extracted from each autocorrelation curve by comparison with two reference states: non-native PepQ-24TMR bound to GroEL and native PepQ-24TMR. The normalized autocorrelation curves of these two reference states are shown in Fig. 2a. The rate of spontaneous PepQ-24TMR folding, measured at 2 nM by FCS, closely recapitulates the rate of folding of the protein observed at 100 nM (Fig. 2b). More importantly, when the same experiment was conducted with fully cycling GroEL-GroES, folding of PepQ-24TMR was stimulated by the same 15–20-fold observed at higher concentrations (Fig. 2b).

Using single-molecule, two-colour co-incidence detection we next probed the assembly status of PepQ during spontaneous folding at 2 nM. As a control, we first examined formation of the native PepQ dimer. A 1:1 mixture of PepQ-24OG and PepQ-24TMR was denatured and refolded at a total PepQ concentration of 100 nM in the presence of the active GroEL-GroES system, in order to permit formation of PepQ dimers carrying both probes. This sample was then diluted to 100 pM PepQ and fluorescence bursts were collected using using a two-channel, confocal-type single-molecule microscope (Fig. 3a, inset). The native PepQ dimer was readily detectable as a robust fraction of coincident events (Fig. 3a). Notably, the observed coincident fraction (~10%) was lower than the theoretically expected value of ~50% for a 1:1 mixture of PepQ-24OG and PepQ-24TMR. This difference is most likely due to the much greater tendency of OG to convert to a long-lived dark (triplet) state, relative to TMR (Supplementary Fig. 2D–F), which results in a substantial decrease in observed co-incidence.

To examine PepQ monomer assembly during spontaneous folding, samples of PepQ-24OG and PepQ-24TMR were mixed at 1:1, denatured in acid-urea and then rapidly diluted (50-fold) into refolding buffer at 23 °C to initiate spontaneous folding at a final protein concentration of 2 nM monomer. This sample was incubated at 23 °C for 10 min, then diluted another 20-fold to a final PepQ monomer concentration of 100 pM. The native PepQ dimer does not readily form at a monomer concentration of 2 nM. However, it is possible that low-order, non-native aggregates stabilized by much larger contact surfaces might still form[36]. We therefore anticipated that any co-incidence observed between the two labelled PepQ monomers would have to result from such low-order aggregates. Importantly, the observed co-incidence was less than 1% (Fig. 3b). Even taking into account the reduced sensitivity caused by the differences in triplet state conversion of the OG and TMR dyes, these measurements indicate that, at most, 4–5% of the PepQ monomers could be found in an assembled state of any kind, including the smallest possible aggregates (non-native dimers), during spontaneous folding at 2 nM. In total, these observations demonstrate that slow spontaneous folding of PepQ cannot be due to inhibitory aggregation, but instead must result from the inherently inefficient conformational search of the PepQ monomer. In

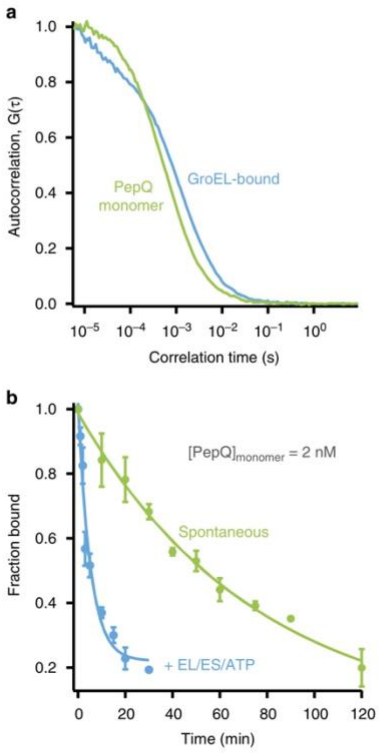


Figure 2 | GroEL accelerates folding of PepQ at very low protein concentrations. (**a**) The difference in diffusion time of the PepQ monomer in free solution, versus bound to a GroEL tetradecamer, can be detected by FCS. The observed FCS curves of the PepQ-24TMR monomer (2 nM), either alone (green) or bound to GroEL (1 μM; blue) are shown. (**b**) Refolding of PepQ-24TMR was monitored by FCS, using the observed shift in diffusion time shown in **a**. PepQ-24TMR was denatured in acid-urea and diluted either directly into buffer (2 nM; spontaneous, green) or into buffer containing wild type GroEL (1 μM). Refolding with GroEL was initiated by addition of GroES (2 μM) and ATP (2 mM; +EL/ES/ATP, blue). At the indicated times, GroEL-mediated folding was quenched by depletion of ATP before FCS measurement, while samples of the spontaneous reaction were mixed with GroEL alone (1 μM) before FCS measurement in order to quench folding and shift the diffusion time of any uncommitted PepQ monomer. The observed fractional change in diffusion time was fit to a single-exponential rate law (solid lines), resulting in rate constants of $0.19 \pm 0.04\ min^{-1}$ for GroEL-mediated folding and $0.013 \pm 0.002\ min^{-1}$ for spontaneous folding. Error bars show the s.d. of three experimental replicates.

addition, our data suggest that the PepQ monomers that do not reach the native state during spontaneous folding at low protein concentrations likely persist as kinetically trapped monomers.

GroEL alters the folding trajectory of the PepQ monomer. To achieve the large folding stimulation observed with PepQ, in the absence of aggregation, GroEL must actively alter how the protein folds. To investigate the nature of this alteration, we exploited the intrinsic tryptophan fluorescence of PepQ. Importantly, PepQ has multiple tryptophan residues, while GroEL and GroES are devoid of this amino acid. During spontaneous folding, the tryptophan fluorescence of PepQ displays a single, downward transition with a time constant of ~125 s (Fig. 4a). The rate of this fluorescence decrease is substantially faster than the limiting rate at which PepQ spontaneously commits to the native state (Fig. 1b). This suggest that, at least for spontaneous folding, the observed shifts in tryptophan fluorescence report on transitions that precede the committed step of PepQ folding. By contrast, assisted folding of PepQ with the cycling GroEL-GroES system results in a rapid, early increase in tryptophan fluorescence ($\tau = \sim 13$ s), which is followed by a subsequent decrease in fluorescence with a time constant of ~73 s (Fig. 4b). The large increase in fluorescence observed with GroEL most likely reports on an early folding transition that occurs after the PepQ folding intermediate has been released into the GroEL-GroES cavity. It is unlikely that the early fluorescence rise is due to either GroES binding and encapsulation alone, or to simple release of the PepQ monomer into the cavity, as these events occur much faster than the observed rate of the PepQ fluorescence change[9,19,37]. Although PepQ folding with the cycling GroEL-GroES system rapidly becomes asynchronous, the transition between the increasing and decreasing fluorescence phases occurs after roughly one cavity lifetime at 23 °C (refs 22,37,38). This observation supports the idea that the increase in fluorescence occurs inside the GroEL-ES cavity. To directly test this conclusion, we employed SR1 to examine a single round of PepQ encapsulation and folding inside the GroEL-GroES cavity. Notably, PepQ confined within the SR1-GroES cavity also displays a rapid increase in fluorescence, but no subsequent decrease (Fig. 4c), confirming the conclusion that the early increase in PepQ fluorescence occurs within the GroEL-GroES cavity.

To further define how the folding behaviour of PepQ is altered by GroEL, we examined the impact of the GroEL C-termini on PepQ folding. We previously showed that a tailless GroEL variant (Δ526 GroEL) has a significantly reduced ability to assist the folding of the classically stringent GroEL-substrate protein, RuBisCO from *R. rubrum*[39]. Interestingly, removal of the C-termini has an even more pronounced negative impact on PepQ folding. Deletion of the C-termini from a cycling GroEL tetradecamer (Δ526) causes a nearly sixfold reduction in the observed PepQ folding rate (Fig. 5a), versus an approximate twofold reduction with RuBisCO (ref. 39). By contrast, removal of the C-termini from the non-cycling SR1 GroEL variant (SRΔ526) results in a more modest twofold decrease in the PepQ folding rate (Fig. 5a). Because deletion of the C-termini can, in some cases, result in premature substrate protein release before GroES binding and encapsulation[40], we considered whether the observed drop in PepQ folding rate with the Δ526 variants is simply due to a trivial decrease in encapsulation efficiency. However, Δ526 displays no substantial premature release of PepQ relative to full-length GroEL and the early escape of PepQ from SRΔ526, compared to SR1, is no greater than 10% (Supplementary Fig. 3). While consistent with previous observations with RuBisCO (ref. 40), this minor drop in encapsulation efficiency is too small to explain the reduction in observed folding rate.

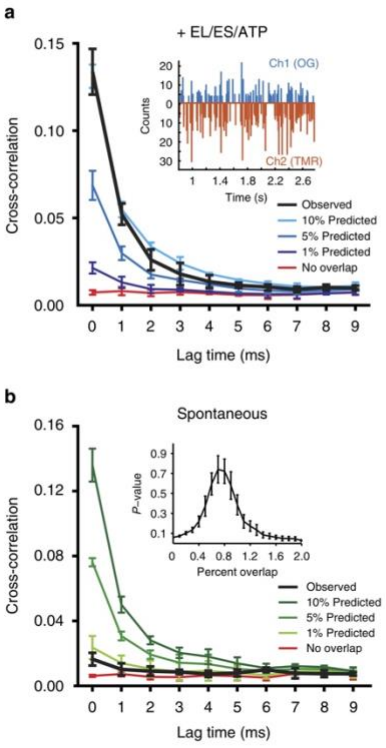


Figure 3 | Non-native PepQ does not aggregate at very low concentrations. (**a**) The formation of the native PepQ dimer was examined with single-molecule, two-colour co-incidence detection. Samples of PepQ-24TMR and PepQ-24OG were denatured, mixed at a stoichiometry of 1:1 (100 nM total PepQ monomer), and then refolded with GroEL, GroES and ATP for 20 min. The sample was diluted 1,000-fold and examined for fluorescence bursts. Examples of the photon history from each detection channel are shown in the inset. Fluorescence burst co-incidence was examined by cross-correlation analysis of the experimental burst data (black). The cross-correlation of numerically generated burst data with known levels of co-incidence (10%, 5%, 1% and no overlap) are also shown. (**b**) Two-colour co-incidence analysis of spontaneous PepQ folding. Samples of PepQ-24TMR and PepQ-24OG were denatured, mixed at a stoichiometry of 1:1 and directly diluted in buffer to a final monomer concentration of 2 nM. The protein was allowed to fold spontaneously at 23 °C for 10 min. The sample was then diluted 20-fold and examined for fluorescence bursts. Cross-correlation analysis of the experimental data set (black), in comparison with numerically generated burst data at known co-incidence levels, are shown. The inset illustrates the *P* value distribution for fitting of the experimental data to numerically generated data sets of known co-incidence, yielding a maximum co-incidence likelihood of <1%.

The stimulation of PepQ folding by the GroEL C-termini could, in principle, result from: (1) enhanced unfolding of PepQ by the C-termini[19,22,23,39]; (2) stimulation of productive folding transitions, or blockage of inhibitory ones, by the tails during intra-cavity folding[20,34,41]; or (3) a combination of both unfolding and confinement effects. Importantly, these models all predict

that the folding trajectory of a PepQ monomer inside the GroEL-GroES cavity should change upon C-terminal tail removal. To test this prediction, we exploited the tryptophan fluorescence properties of PepQ to examine a single round of encapsulation inside both the full-length SR1-GroES cavity and the truncated SRΔ526-GroES cavity. Strikingly, the rapid and early rise in tryptophan fluorescence that is observed when PepQ folds inside the SR1-GroES cavity, completely disappears when PepQ is encapsulated inside a SRΔ526-GroES cavity (Fig. 5b). These observations are highly consistent with the idea that GroEL promotes conformations of the PepQ monomer that are not, or at least not well, populated during spontaneous folding in free solution and that the C-terminal tails are at least partially involved in this process.

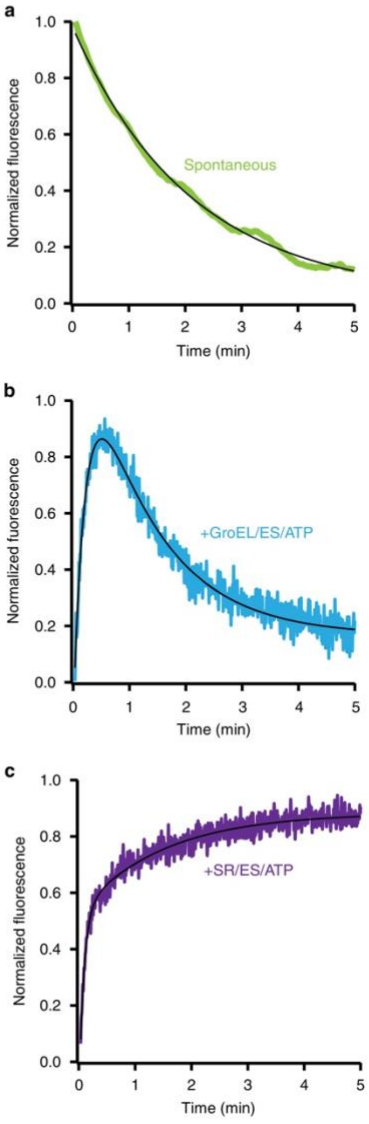


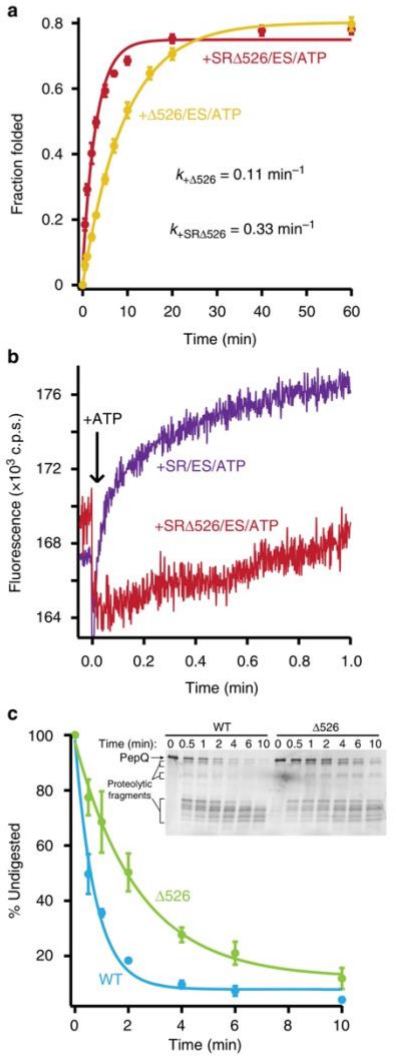
Figure 4 | GroEL alters the folding trajectory of the PepQ monomer. Folding of PepQ was monitored by changes in intrinsic tryptophan fluorescence during (**a**) spontaneous folding, (**b**) folding by the fully cycling GroEL-GroES system (**c**) folding after a single round of encapsulation within the SR1-GroES complex. For spontaneous folding, wild-type PepQ was first denatured in acid-urea then diluted directly into buffer (100 nM). For GroEL-GroES folding, acid-urea denatured PepQ (100 nM) was bound to wild-type GroEL (200 nM) and refolded in the presence of GroES (400 nM) and ATP (2 mM). For SR1-GroES folding, acid-urea denatured PepQ (100 nM) was bound to SR1 (300 nM) and refolded in the presence of GroES (600 nM) and ATP (2 mM). In all cases, the traces shown represent the average of 10 independent experimental replicates. All traces were fit (solid lines) to either a single-exponential rate law (spontaneous) or a sum of exponentials (GroEL-GroES and SR1-GroES). The observed rate constants were $-0.477 \pm 0.003\,min^{-1}$ for spontaneous folding, $4.63 \pm 0.05\,min^{-1}$ and $-0.826 \pm 0.007\,min^{-1}$ for GroEL-GroES folding and $11.5 \pm 0.2\,min^{-1}$ and $0.669 \pm 0.010\,min^{-1}$ for SR1-GroES folding.

Cryo-EM observation of PepQ unfolding by the GroEL C-termini. We previously demonstrated that GroEL helps stimulate productive folding of a kinetically trapped RuBisCO monomer through partial unfolding[19,22,23,39]. In addition, we showed that maximal RuBisCO unfolding requires the GroEL C-terminal tails[40]. If structural disruption of the misfolded substrate proteins is a general feature of GroEL-stimulated folding, then GroEL could also be expected to unfold the kinetically trapped PepQ monomer. To test this proposition, we first examined the protease susceptibility of a PepQ folding intermediate bound to both wild-type GroEL and Δ526. Chemically denatured PepQ was first bound to the open, *trans* ring of an asymmetric GroEL-GroES complex created with either wild-type GroEL or Δ526, then treated with limiting amounts of chymotrypsin[23]. Consistent with our previous RuBisCO observations, PepQ bound to a full-length GroEL ring was degraded ~2.5-fold faster than PepQ bound to the Δ526 ring (Fig. 5c).

To develop a more detailed picture of the interaction between PepQ and GroEL, we employed cryo-EM to examine the structures of both wild type GroEL and Δ526 tetradecamers bound to non-native PepQ. Chemically denatured PepQ was first mixed with unliganded (apo) GroEL or Δ526 tetradecamers, then vitrified in thin ice and imaged with single-particle cryo-EM. Reference-free two-dimensional (2D) image classification revealed a robust population of GroEL tetradecamer complexes with substantial density visible in the central cavity of a major 2D class-average for both wild-type GroEL and Δ526 (Supplementary Fig. 4). The observed central density is highly consistent with the expected binding position of the non-native PepQ monomer. Further, three-dimensional (3D) classification and map refinement, without any applied symmetry, revealed both apo and PepQ-bound states of the tetradecamer for both wild-type GroEL and Δ526. (Fig. 6 and Supplementary Figs 5 and 6). On the basis of the gold-standard Fourier shell correlation, the overall resolution for the apo states of both wild-type GroEL and Δ526 was 7.9 Å, while the overall resolution of the pepQ-bound states for both wild-type GroEL and Δ526 was 8.3 Å (Supplementary Fig. 7A). Importantly, the resolution obtained was not uniformly distributed over the entire structure of either complex, but was

significantly higher in the equatorial and intermediate domains, and lower in the apical domains (Supplementary Fig. 7B–E). The lowest local resolution was observed for density associated with the PepQ monomer in the central cavity, as expected for a non-native protein folding intermediate, which likely populates a mixture of conformations.

The cryo-EM structures reveal the striking impact of the GroEL C-termini on both the conformation and the binding position of the PepQ folding intermediate. In the absence of the C-terminal tails, the PepQ monomer appears as a strong extra density associated with the upper, inner surface of the apical domains of the GroEL cavity (Figs 6c and 8d). By contrast, on a wild-type GroEL ring, the PepQ folding intermediate shifts to a much lower average position in the cavity, moving towards the base of the cavity and in the direction of the C-termini. At the same time, the density of the PepQ intermediate decreases significantly (Figs 6f and 8c), which indicates a more unfolded and heterogenous conformational ensemble of the PepQ, leading to its weaker density in the cryo-EM map. The non-native PepQ monomer can also be seen to make contact with multiple GroEL subunits on both a wild-type GroEL and Δ526 ring (Fig. 7). However, the location of the contacts between the PepQ folding intermediate and the GroEL subunits changes dramatically when the C-terminal tails are removed. In the tailless Δ526 ring, the PepQ monomer appears to make exclusive contact with the central face of the apical domains, in the region of helices H and I (Figs 7a–c and 8d). By contrast, on a wild-type GroEL ring, the PepQ monomer shifts to a significantly lower position at the base of the apical domain, and is accompanied by a set of new, strong contacts that localize in the region of the GroEL C-terminal tails (Figs 7d–f and 8c). To confirm that the density observed at the base of the wild-type GroEL cavity does, in fact, originate from the C-termini, we examined this region in the empty wild-type GroEL and Δ526 tetradecamers. As expected, the density observed at the base of the wild-type GroEL ring, projecting from the precise position expected for the C-termini, is missing in the tailless Δ526 ring (Supplementary Fig. 8A,B). In total, these result strongly support the idea that the non-native PepQ monomer is significantly more unfolded when bound to a GroEL ring with intact C-terminal tails and this unfolding has a direct functional impact on the efficiency of productive folding.



Discussion

Fundamentally, chaperonins like GroELS function as kinetic editors of protein folding reactions, altering how folding intermediates partition between available conformational states.

Figure 5 | The GroEL C-termini alter the conformation and folding of the PepQ monomer. (**a**) Acid-urea denatured PepQ was bound to a C-terminal truncation mutant of tetradecamer GroEL, Δ526 (200 nM, red) or the single-ring truncation mutant, SRΔ526 (300 nM, orange) and refolded in the presence of GroES (400 and 600 nM, respectively) and ATP (2 mM). In each case, the observed regain in enzymatic activity was fit to a single-exponential rate law (solid lines), resulting in observed rate constants of 0.106 ± 0.003 min^{-1} for Δ526-mediated folding and 0.332 ± 0.038 min^{-1} for SRΔ526-mediated folding. (**b**) Intra-cavity folding of PepQ at early times was monitored by changes in tryptophan fluorescence following addition of GroES and ATP to complexes of non-native PepQ bound to SR1 (blue) or SRΔ526 (green). Acid-urea denatured PepQ (100 nM) was first bound to either SR1 or SRΔ526 (300 nM in both cases), and then rapidly mixed with an equal volume of GroES (600 nM) and ATP (2 mM) in a stopped-flow apparatus. The traces shown represent the average of 20 experimental replicates. (**c**) Residual structure in a GroEL-bound PepQ folding intermediate was examined by protease susceptibility. PepQ-24F (100 nM) was denatured in acid-urea and bound to the *trans* ring of either wild-type GroEL-GroES or Δ526-GroES ADP complexes (ref. 23; 120 nM) and then treated with chymotrypsin for the indicated times before quenching with phenylmethylsulfonyl fluoride (PMSF) (1 mM). Samples were analysed by SDS–PAGE and laser-excited fluorescence gel scanning (inset). The migration position of full-length PepQ, as well as the position of three dominant groups of proteolytic fragments, are indicated. The amount of full-length PepQ was quantified by densitometry. The data were fit to a single-exponential rate law, with a half-time for the digestion of PepQ bound to the open ring of a wild type GroEL-GroES-ADP complex of 0.53 ± 0.06 min (EL, blue) and 1.66 ± 0.17 min for the Δ526-GroES ADP complex (Δ526, green). In all cases, error bars show the s.d. of three experimental replicates.

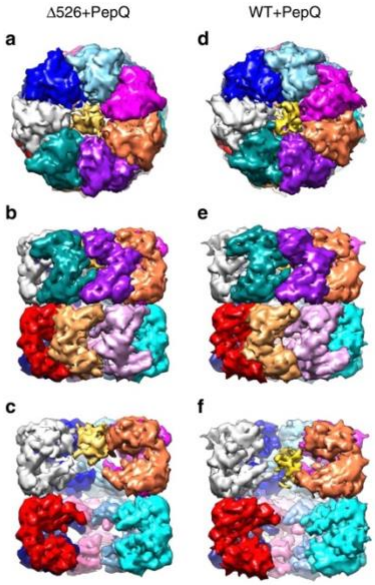


Figure 6 | The impact of the GroEL C-termini on a bound PepQ monomer. The conformation of a non-native PepQ monomer bound to either a wild-type GroEL tetradecamer (*WT*) or to the C-terminal truncation variant (Δ526) was examined by cryo-EM. The map of the Δ526-PepQ complex is shown from the top (**a**), side (**b**) and cutaway (**c**) views. The map of the wild-type GroEL-PepQ complex is also shown from the top (**d**), side (**e**) and cutaway (**f**) views. Each subunit of GroEL is coloured differently while the PepQ monomer is coloured yellow. In the Δ526 GroEL, the density attributed to the PepQ monomer is observed near the top of the GroEL cavity in one ring (**c**). However, in the wild-type GroEL complex, the PepQ density is observed near the centre of the cavity (**f**) in one ring.

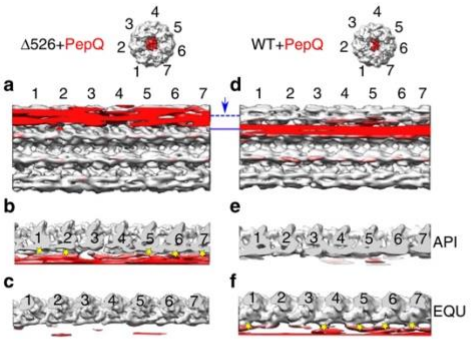


Figure 7 | The PepQ monomer interacts with multiple GroEL subunits. GroEL tetradecamers with bound PepQ are shown in unwrapped, planar displays as viewed from inside the ring for the Δ526 (**a**) and wild-type (**d**) GroEL complexes. A top view slice of the planar map, through the apical (API) domains (dashed blue line), is shown in **b** for Δ526 and in **e** for wild-type GroEL. Interactions between the non-native PepQ monomer and the Δ526 GroEL apical domains of subunits 1, 2, 5, 6 and 7 are highlighted (**b**; yellow stars). A top view slice of the planar map, through the equatorial (EQU) domains (solid blue line), is shown in **c** for Δ526 and in **f** for wild-type GroEL. Contacts between the non-native PepQ and the C-termini of subunits 1, 4, 5, 6 and 7 of the wild-type GroEL are highlighted (**f**; yellow stars). The isosurface threshold for **b** and **c** is 1.74σ and is 1.65σ for **e** and **f**.

A key question, however, is whether chaperonins achieve this editing action by actively altering the conformational space available to their substrate proteins, or by exclusively working as passive aggregation inhibitors. We examined this issue from a new angle by characterizing the folding of the *E. coli* metalloprotease PepQ, a stringent, *in vivo* GroEL-substrate protein. We found that slow spontaneous folding of PepQ is not caused by inhibitory aggregation. The capture of this kinetically trapped PepQ folding intermediate by a GroEL ring results in conformational perturbations that are consistent with unfolding. In addition, the intrinsically unstructured C-terminal tails of the GroEL subunits play a central role in this process (Fig. 9a).

Determining the function of the flexible C-terminal tails in chaperonin-assisted protein folding has been challenging. Early studies showed that the tails play no role in tetradecamer assembly or stability[42]. At the same time, removal of the C-termini was found to have negligible impacts on *E. coli* growth under standard laboratory conditions, leading to the suggestion that the tails do not play any important role in assisted folding[42,43]. Other studies, however, demonstrated that removal of, or large alterations to, the C-termini can have serious negative consequences in *in vitro* protein folding assays[34,39,41,44,45]. In addition, *E. coli* strains possessing C-terminally truncated GroEL genes display substantially compromised fitness in competition with wild-type strains[42]. These observations, in combination with the extensive, although not quite universal, conservation of the chaperonin C-terminal tails over much of phylogeny[46,47] suggest that the C-termini do play an important role in assisted protein folding. Our prior work with RuBisCO supported this conclusion, implicating the C-termini in substrate protein capture, retention and unfolding during GroES binding[39,40]. The observations we present here with PepQ strengthen and extend these conclusions, showing that the unstructured C-termini make physical contact with a non-native substrate protein before ATP or GroES binding. In addition, we have visualized the consequences of this interaction, demonstrating simultaneous engagement of a folding intermediate by both the inner apical face and the unstructured tails of multiple GroEL subunits. This multi-level binding mode both retains the folding intermediate deeper inside the GroEL cavity and assists in partial unfolding of the misfolded PepQ monomer.

Interestingly, we observe a single, well populated class of the PepQ folding intermediate bound to a GroEL ring, both in the presence and absence of the GroEL C-termini. This contrasts with a previous cryo-EM study conducted with the the smaller substrate protein malate dehydrogenase (MDH), in which asymmetric model refinement suggested multiple potential binding modes of the MDH folding intermediate[48]. While the C-termini were not resolved in this prior study, and the resolution of these MDH structures is several angstroms lower than the PepQ structures we report here, two sub-populations of the MDH folding intermediate appear to be bound in a deep internal position within the GroEL cavity, consistent with the binding position we observe with PepQ. A third sub-population of the bound MDH monomer appeared to be bound in a more elevated position near the upper, exterior surface of the GroEL apical domains[48]. At the same time, the MDH folding intermediate displayed a substantially smaller contact surface

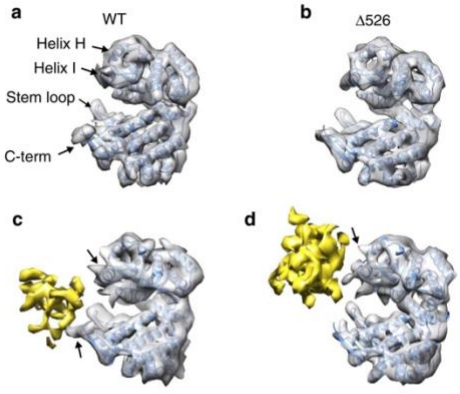


Figure 8 | The GroEL C-termini helps retain and unfold the PepQ monomer. (**a**,**b**) A single subunit of the apo GroEL atomic model (PDB ID: 4HEL) fit into the cryo-EM densities of unoccupied wild-type GroEL and Δ526 GroEL tetradecamers. The positions of the H and I helices of the GroEL apical domain, as well as the equatorial stem loop (D41-P47) and the C-terminus are labelled on the wild type GroEL structure. (**c**,**d**) Single subunit of the apo GroEL atomic model fit into the cryo-EM densities of PepQ-bound wild-type GroEL and Δ526 GroEL tetradecamers. The density from the non-native PepQ monomer is coloured yellow. Black arrows (**c**,**d**) indicate the interactions between a GroEL subunit and the PepQ monomer. When the density volumes of the Δ526 and wild-type GroEL tetradecamers are matched (~61,000 Å^3 for a single GroEL subunit in both cases), the observed density volumes for the PepQ monomer are 8,564 Å^3 in the Δ526 and 2,696 Å^3 in the wild type GroEL complex, consistent with the PepQ monomer being more unfolded when bound to the wild type GroEL ring.

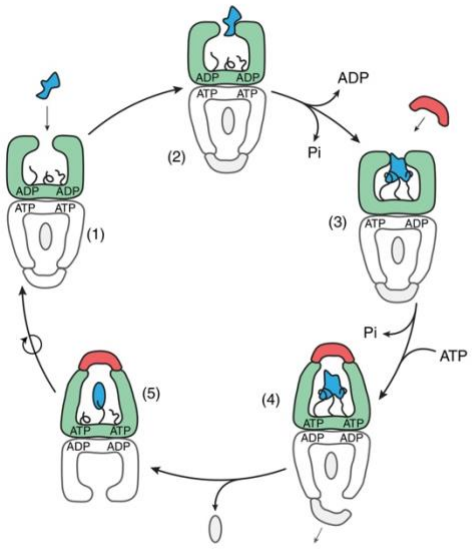


Figure 9 | Schematic of the GroEL-GroES reaction cycle. (1) A non-native substrate protein (irregular blue shape) enters the GroEL reaction cycle on the open *trans* ring (green) of the ATP bullet complex[22]. (2) Substrate protein binding accelerates both the release of ADP from the *trans* ring and ATP hydrolysis in the opposite, *cis* ring (grey; refs 50,51). (3) Binding of the non-native substrate protein by the C-terminal tails (black), helps retain the substrate protein deep within the GroEL cavity and, in combination with additional binding by multiple apical domains, results in substrate protein unfolding (results here and refs 19,22,24,39,40,55). (4) Assembly of the new folding cavity on the *trans* ring causes both forced unfolding and compaction of the substrate protein, and is directly coupled to the disassembly of the folding cavity on the opposite ring, potentially through a transient, symmetric intermediate[22,23,37,77–79]. (5) A subsequent allosteric shift of the GroEL-GroES complex results in full ejection of the substrate protein into the enclosed GroEL-GroES cavity and the initiation of folding before ATP hydrolysis[50]. Because ATP hydrolysis is the rate limiting step of the reaction cycle, increased binding of substrate proteins to the open *trans* ring (2) results in more rapid cycling of the GroEL-GroES system and a shorter folding cavity lifetime[22,50,51,80].

with the GroEL ring[48] in comparison to what we observe with PepQ (Fig. 7). This observation is consistent with the difference in relative mass of MDH (33 kDa) compared to PepQ (52 kDa) and suggests that the smaller MDH protein could be bound more weakly or sample a larger potential range of bound states.

In principle, the cryo-EM structure of the PepQ-GroEL complex could also reveal conformational changes of the GroEL tetradecamer that are coupled to substrate protein capture. Overall, the conformations of the GroEL tetradecamer in the presence and absence of the PepQ folding intermediate are similar. Notable breaks in the rotational symmetry of the PepQ-occupied GroEL ring, both in the presence and absence of the C-termini, are apparent (Fig. 7). However, significant deviations from ideal rotational symmetry are also observed in the unoccupied apo GroEL and Δ526 rings (Supplementary Fig. 8C–J). Strikingly, PepQ binding induces a dramatic increase in the rotational symmetry of the Δ526 apical domains, in both the bound and second, unoccupied rings (Supplementary Fig. 8H,J). The structural changes seen in the second, unoccupied ring are most likely a consequence of allosteric coupling between the GroEL rings. The coordinated binding and release of nucleotides, GroES and substrate proteins are well established and essential features of the functional GroEL reaction cycle[49]. Many of the structural details of this allosteric coupling remain poorly understood, however. In particular, it remains unclear how substrate protein binding forces ADP out of one GroEL ring while simultaneously accelerating the release of GroES from the opposite ring[37,50,51]. Previous work suggested that this allosteric response may involve counter-clockwise movements of the GroEL apical domains, in both the substrate occupied and second, empty ring[52]. Our structural observations with PepQ suggest that a shift in the rotational symmetry of the GroEL apical domains likely also plays a role. In addition, the C-termini appear to be intimately involved in modulating this structural shift. This observation is consistent with our prior observations that removal of the C-termini attenuates negative cooperativity in ATP binding between the two GroEL rings[39].

Overall, our observations with PepQ are not consistent with an exclusively passive, aggregation-blocking role for GroEL in stimulated protein folding[53,54]. These observations are, however, fully consistent with our previous demonstration that GroEL plays an active role in the assisted folding of *R. rubrum* RuBisCO (refs 19,22,23,39). They are also consistent with observations from other groups on other substrate proteins, including another endogenous *E. coli* enzyme DapA[20,21,24,25,55,56]. Our observations

with PepQ also suggest that active participation by GroELS in stimulated protein folding is likely to be a general mechanistic feature of these chaperonin machines. DapA, like RuBisCO, is a member of the TIM-barrel family of proteins, a canonical α/β-fold that is highly represented in the subset of *E. coli* proteins that depend on GroEL for folding[31,32]. By contrast, PepQ is a member the so-called pita-bread proteins[28–30], a protein fold that is fundamentally distinct from the TIM-barrel fold[27]. To date, no pita-bread fold has been examined in detail as a GroEL-substrate protein. The addition of PepQ to the list of *E. coli* proteins that derive a large, active folding enhancement from GroEL strengthens the argument that similar mechanisms are likely to stimulate the folding of many stringent substrate proteins.

Interestingly, PepQ appears to have no ready access to fast and productive folding pathways in free solution. At the same time, persistent misfolding produces PepQ monomers that, although they do not aggregate, cannot reach the native state even with assistance from GroEL. This suggests that the conformational search of the non-native PepQ monomer, at least in free solution, is dominated by deep and inhibitory kinetic wells that GroEL helps the protein to avoid. Whether the iterative annealing or confinement-based models most accurately describe this active folding mechanism of GroEL remains controversial[5]. Importantly, these mechanisms make distinct predictions about what should happen to PepQ folding when the GroEL cycling rate is altered. If unfolding of kinetically trapped intermediates is important for stimulated folding of PepQ, it should be possible for the cycling GroEL-GroES system to achieve a stimulated folding rate that exceeds the limiting, intra-cavity folding rate observed with SR1-GroES. By contrast, if confinement is most important for PepQ folding, then the non-cycling SR1-GroES system should display the maximum possible enhanced folding rate, a rate that the cycling system could approach but never exceed[22].

To test these predictions, we examined the folding rate of PepQ under conditions where the GroEL-GroES cycling rate was systematically increased. Modulation of the GroEL ATPase rate was accomplished by addition of bovine serum albumin (BSA), which interacts only weakly with GroEL. Because progression of the GroEL ATPase cycle is linked to ADP release, which is in turn coupled to binding of proteins to the open, post-hydrolysis *trans* GroEL ring (Fig. 9 and refs 37,50,51), BSA can be used to accelerate the GroEL-GroES ATPase cycle (Fig. 10a). However, because the interaction between BSA and GroEL is weak, BSA only poorly competes with PepQ for binding to GroEL. At concentrations up to 0.1 mg ml^{-1}, BSA has a small, negative impact on the observed PepQ folding rate observed with cycling GroEL-GroES (Fig. 10b). Strikingly, however, while the addition of BSA has no impact on either spontaneous PepQ folding or SR1-GroES mediated folding, higher BSA concentrations substantially enhance the PepQ folding rate achieved with cycling GroEL-GroES (Fig. 10b). Importantly, the magnitude of this effect increases as the concentration of BSA increases, mirroring the impact of BSA on the steady-state rate of ATP turnover by GroEL-GroES (Fig. 10b). This response is very similar to our prior observations with RuBisCO (ref. 22), where an ~40% increase in the steady-state GroEL-GroES ATPase rate yielded a 2.5–3-fold enhancement of the observed RuBisCO folding rate. These observations suggest that repetitive unfolding by GroEL is required to achieve maximally stimulated folding of many stringent substrate proteins.

When considering the stimulatory impact of partial unfolding, it is important to note that GroEL unfolds substrate protein in two distinct phases. The first is associated with the capture of a folding intermediates by the GroEL ring, where a binding-driven expansion of the substrate protein can, in some cases, result is substantial conformational disruption (this study and refs 19,22–24,39,55). This unfolding event occurs both during and immediately after capture of a folding intermediate, but before ATP binding. Most likely, binding-associated unfolding is similar to surface-catalysed denaturation, where the substrate protein becomes splayed across the multiple interaction surfaces of the apical domains as well as, we suggest, the C-termini[19,22,23,39,55].

GroEL also imposes a second, directed unfolding process that is impelled by ATP[22–24,39]. When a GroEL ring binds ATP, the apical domains are driven through a large-scale, rigid body rearrangement that both rotates and elevates them[57–59]. While these shifts are necessary for GroES binding and substrate encapsulation[9–12], previous observations have also demonstrated that (1) substrate proteins remain associated with the apical

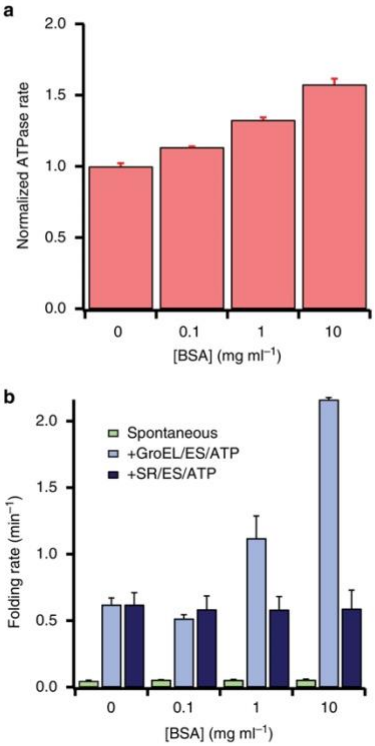


Figure 10 | Cycling GroEL-GroES can fold PepQ faster than confinement alone. (**a**) The rate of ATP hydrolysis by GroEL-GroES is stimulated in the presence of BSA. The steady-state rate of ATP hydrolysis by GroEL (200 nM) in the presence of GroES (400 nM) and ATP (2 mM) was measured with varying concentrations of [BSA]. Error bars show the standard deviation of three independent experiments. (**b**) Addition of BSA to a cycling GroEL-GroES system substantially accelerates the rate of assisted PepQ folding. The rate of spontaneous PepQ folding (green), intra-cavity folding with SR1-GroES (dark blue), and folding with fully the cycling wild type GroEL-GroES system (light blue) was examined in the presence of different concentrations of BSA. Experimental conditions were the same as Fig. 1, with the exception that native BSA was present in the buffer. Error bars show the s.d. of three experimental replicates.

domains as they initiate their movement, imposing a substantial load on their motion[60] and (2) apical domain movement can simultaneously impart a rapid, forced unfolding event on the substrate protein[22–24,39]. While our studies with PepQ were not designed to detect forced unfolding, it is striking that both binding-driven unfolding of PepQ and RuBisCO (this study and ref. 39) and forced unfolding of RuBisCO[39] are attenuated when the C-termini are removed. These observations suggest that the C-termini represent a secondary binding platform at the base of the GroEL cavity that is important both for the initial capture and unfolding of the substrate protein, as well as retention of the folding intermediate within the GroEL cavity during the process of apical domain movement and GroES binding[40]. It remains an open question how the C-termini are induced to release the substrate protein upon the initiation of folding. However, both experimental[39] and computational studies[61] indicate that the C-termini are coupled to the GroELS allosteric cycle, suggesting that modulation of the interaction between the C-termini and a folding intermediate might be controlled by the GroEL ATPase cycle in a manner that parallels the behaviour of the apical domains.

Fundamentally, the iterative annealing and confinement mechanisms are not mutually exclusive. A combined mechanism, in which kinetically trapped folding intermediates are first partially unfolded, then briefly confined within the privileged environment of the GroEL-GroES cavity where re-population of misfolded conformations is discouraged, might well yield a maximally efficient strategy for accelerating the folding of especially recalcitrant proteins. Several of our observations with PepQ are consistent with such a mechanism. In the presence of either single-ring or double-ring GroEL variants, PepQ displays a sizable fluorescent burst phase that is completely absent during spontaneous folding (Fig. 4). This observation suggests that the PepQ monomer, while confined within the GroEL-GroES cavity, populates at least one conformational state (or ensemble of states) with ready access to the native state. During spontaneous folding; however, this state is either very rarely populated, or not populated at all. At the same time, removal of the GroEL C-terminal tails slows overall PepQ folding and completely eliminates the fluorescence burst phase (Fig. 5). This behaviour is strikingly similar to the impact of C-terminal tail removal on RuBisCO folding, where the formation of a rapidly folding, burst phase intermediate depends upon both partial unfolding and encapsulation within the GroEL-GroES cavity[23,39,40]. As with RuBisCO, C-terminal tail removal also has a more profound impact on PepQ folding with the cycling, tailless Δ526 tetradecamer than it does on the tailless single-ring SRΔ526 (Figs 1 and 5). For both RuBisCO and PepQ, however, long-term confinement within the chaperonin cavity, even when partial unfolding is reduced through C-terminal tail removal (for example, SR1-GroES versus SRΔ526-GroES) results in substantially enhanced folding in comparison to the free solution folding of both proteins (Figs 1 and 5 and refs 3,20,39). In total, these observations are consistent with an active chaperonin mechanism in which partial unfolding and confinement lead to optimal stimulation of folding for highly dependent substrate proteins. It is worth noting that in a living *E. coli* cell, additional chaperone systems (for example, the Hsp70s and Hsp100s) can engage a folding intermediate before its processing by GroELS[1]. Learning how these additional chaperone systems impact the folding of GroELS substrates will be an important next step towards understanding the mechanism of chaperone and chaperonin-mediated folding pathways.

Methods

Bacterial strains. All bacterial strains used in this work were originally obtained from the laboratory of Dr. Arthur Horwich at Yale University Medical School.

Bacterial strains employed:
BL21—*E. coli* B dcm ompT hsdS(rB-mB-) gal
BL21DE3—*E. coli* B dcm ompT hsdS(rB-mB-) gal [λDE3]
DH5α—*E. coli* fhuA2 lac(del)U169 phoA glnV44 Φ80' lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17.

Proteins. Wild type and variants of GroEL (SR1 and C-terminal truncation mutants), GroES and wild type *E. coli* PepQ were all expressed and purified as described previously[19,22,23,39,50]. The cysteine mutant of PepQ, A24C, was generated via standard site-directed mutagenesis[62] and the sequence was verified by DNA sequencing. This mutant was expressed and purified following the protocol for wild type PepQ.

In brief, GroEL was expressed from an inducible plasmid in *E. coli* BL21 in LB at 37 °C. After cell disruption, the crude lysate was clarified by ultracentrifugation (142,000 × *g*), followed by anion exchange chromatography (FastFlow Q, GE) at pH 7.4. GroEL fractions were concentrated by 70% (w/v) ammonium sulfate precipitation. This precipitate was solubilized and dialyzed against buffer at pH 6.8 containing 25% (wild-type GroEL) or 12.5% (all GroEL mutants) methanol. A second round of strong anion exchange (FastFlow Q, GE), run in the same methanol-containing buffer at pH 6.8, was used to strip co-purifying small proteins and peptides from the GroEL oligomers. To further remove contaminating proteins and peptides that remain tightly associated through prior stages of purification, GroEL fractions were gently agitated in the same methanol-containing buffer and Affi Blue Gel resin overnight at 4 °C under an argon atmosphere. The final sample was dialysed into storage buffer (pH 7.4), supplemented with glycerol (15–20% v/v), concentrated, and snap frozen using liquid N_2.

GroES was expressed from an inducible plasmid in *E. coli* BL21(DE3) in LB at 37 °C. After cell disruption, the crude lysate was clarified by ultracentrifugation (142,000 × *g*), followed by acidification with sodium acetate, and anion exchange chromatography at pH 4.6 (FastFlow Q, GE). The sample was dialysed against buffer at pH 7.4 and applied to a strong anion exchange column (Source Q, GE). GroES was eluted with NaCl and enriched fractions were pooled. The sample was dialysed into storage buffer (pH 7.4), supplemented with glycerol (15–20% v/v), concentrated and snap frozen using liquid N_2.

PepQ and PepQ mutants were expressed from an inducible plasmid in *E. coli* BL21(DE3) in LB at 37 °C. After cell disruption, the crude lysate was clarified by ultracentrifugation. The supernatant was applied to a strong anion exchange column (FastFlow Q, GE) at pH 7.4 and eluted with a gradient of NaCl. Fractions enriched for PepQ were pooled, and the protein was precipitated with 70% (w/v) ammonium sulfate. The sample was loaded on a hydrophobic interaction column (Phenyl Sepharose FF, GE) at pH 7.4 and eluted with a decreasing ammonium sulfate gradient. Fractions enriched for PepQ were pooled, dialysed into storage buffer (pH 7.4), supplemented with glycerol (15–20% v/v), concentrated and snap frozen using liquid N_2.

Labelling of PepQ. A24C PepQ was labelled using either 5-iodoacetamido-fluorescein (fluorescein, F), 5-(2-acetamidoethyl) aminonaphthalene 1-sulfonate (EDANS, ED), tetramethylrhodamine-5-iodoacetamide dihydroiodide (tetramethylrhoadmine, TMR), or Oregon Green 488 iodoacetamide (Oregon Green, OG). All dyes were obtained from Invitrogen (Molecular Probes). PepQ (~10 mg ml^{-1} in 50 mM Tris, pH 8, 100 mM KCl, 1 mM $MgCl_2$) was reduced with 0.5 mM tris(2-carboxyethyl)phosphine (TCEP) TCEP and labelled with a 10-fold excess of reactive dye, added in 1 addition for 3 h at 23°C. The reaction was quenched by adding glutathione (5 mM), and the labelled PepQ was first separated from unbound dye by gel filtration (PD-10 column, Pharmacia), followed by re-purification of the labelled protein with high-resolution ion exchange chromatography (MonoQ, GE). The extent of labelling was determined by protein quantification by the Bradford assay (Bio-Rad) and dye quantification under denaturing conditions using known dye molar extinction coefficients[37,63]. Unique labelling of a single cysteine was verified by both denaturing anion exchange chromatography (MonoQ, GE) in 8 M urea buffer and by detection of a single major and fluorescent tryptic peptide upon separation by C8 reverse-phase chromatography[63].

Folding assays. PepQ refolding assays were conducted in TKM buffer (50 mM Tris-HCl, pH 7.4, 50 mM KOAc, 10 mM $Mg(OAc)_2$ and 2 mM DTT) using a protocol similar to that employed previously for RuBisCO (refs 19,22,23,37,50,63), with differences in the folding buffer composition, duration of post-reaction incubation, and the detailed assay method[27]. All folding assays were conducted using PepQ that was diluted at least 40-fold into 8 M urea, 25 mM glycine phosphate, pH 2, and incubated at room temperature for at least 20 min before further use. CD spectra show a complete loss of secondary structure under these conditions (Supplementary Fig. 9). Spontaneous refolding of PepQ was initiated by a 50-fold dilution from denaturant into TKM buffer (50 mM Tris-HCl, pH 7.4, 50 mM KOAc, 10 mM $Mg(OAc)_2$ and 2 mM DTT) and quenched through the addition of excess GroEL. Chaperonin-mediated folding reactions using either wild-type or mutant tetradecameric GroEL began with a 50-fold dilution of denatured PepQ into TKM buffer containing a particular GroEL variant. GroES and ATP were added to initiate the reaction cycle and the reaction was quenched

with hexokinase and glucose[22,23,39]. Folding reactions in single-ring mutants of GroEL were done similarly, except quenching was accomplished by the simultaneous addition of EDTA and incubation of the sample at 0 °C (refs 20,34). After quenching, all samples were incubated for 60 min at room temperature to allow for dimerization. The enzyme activity of all samples was measured through an NAD-coupled reaction using alanine dehydrogenase from *B. subtilis*[27].

Measuring PepQ persistence in solution. Fluorescein-labelled PepQ (24F) was allowed to refold spontaneously or in the presence of the chaperone system (as in Fig. 1b, see Folding assays in Methods section). Samples were taken after 60 min and run on 10% SDS–polyacrylamide gel electrophoresis (SDS–PAGE). Gels were imaged with a Typhoon Trio (GE Healthcare) and quantified with ImageJ.

Fluorescence and light scattering. Light scattering and fluorescence measurements were conducted with a T-format fluorometer (PTI), equipped with a jacketed cuvette holder for temperature control with a high-precision circulating water bath (Neslab). For both types of experiments, the assays were initiated by diluting acid-urea denatured PepQ at least 50-fold into temperature-equilibrated TKM buffer (23 °C) in the presence or absence of GroEL. Tryptophan fluorescence was monitored with excitation at 295 ± 4 nm and emission at 340 ± 4 nm. The excitation and emission wavelengths were both 340 ± 1 nm for light scattering experiments.

Stopped-flow fluorescence. Stopped-flow experiments were conducted using an SFM-400 rapid mixing unit (BioLogic) equipped with a custom-designed two-channel fluorescence detection system[19,23,39,63]. Mixing was done using equal volume injections from two syringes, one containing GroEL-PepQ complexes and one containing GroES and ATP. Each solution was in TKM buffer. Measurements were taken every 150 ms.

Steady-state FRET. Steady-state fluorescence measurements were conducted with a with a T-format fluorometer (PTI), equipped with a jacketed cuvette holder for temperature control with a high-precision circulating water bath (Neslab). FRET efficiencies were calculated from the changes in donor-side fluorescence of matched donor only and donor plus acceptor labelled molecules[37,63].

Single-molecule fluorescence microscope. Built on a research quality, vibrationally isolated 4′ × 8′ optical table, the system is constructed around a Nikon Eclipse Ti-U inverted microscope base using a ×60/1.4NA CFI Plan Fluor oil immersion objective. The microscope base is outfitted with a precision, 2-axis stepper motor sample stage (Optiscan II; Prior) and a custom-designed confocal optical bench with three, independent detection channels. Each detection channel is configured with an optimized band-pass filter set for wavelength selection and a low-noise, single photon counting APD unit (SPCM-AQRH-15; Excelitas). Photon pulses are collected and time stamped with either a multichannel hardware correlator (correlator.com) or high speed TTL counting board (NI9402; National Instruments). Sample excitation is provided by either one or a combination of three lasers: two diode lasers (488 and 642 nm; Omicron) and one diode-pumped solid state laser (561 nm; Lasos). The free-space beams of each laser are each coupled to a three-channel fibre combiner (PSK-000843; Gould Technologies) and the combined output is directed into the sample objective with a custom, triple-window dichroic filter (Chroma). Each laser is addressable from the integrated control and data acquisition software, custom developed using LabView (National Instruments).

PepQ refolding by fluorescence correlation spectroscopy. PepQ-24TMR was diluted greater than 40-fold (to 5 μM) into 8 M urea, 25 mM glycine phosphate, pH 2 and incubated for 20 min at room temperature. For spontaneous folding reactions, this PepQ-TMR was then diluted to 100 nM in the same solution. The folding reaction was initiated by dilution of PepQ to 2 nM in TKM buffer. Folding was quenched by the addition of 50 μl of the refolding reaction to 50 μl of 1 μM GroEL in TKM buffer. For GroEL-mediated folding, 5 μM denatured PepQ-24TMR was diluted to 100 nM in TKM buffer containing GroEL (200 nM final tetradecamer concentration). After a 10 min incubation at room temperature, this solution was diluted into TKM buffer containing GroEL, GroES and an ATP-regeneration system[22]. Folding was then initiated by the addition of ATP. The final concentration of ATP was 2 mM, GroEL was 1 μM, and GroES was 2 μM. Folding was quenched by the addition of 20 μl of the reaction mixture with an equal volume of hexokinase and glucose. Dimerization was not observable in refolding assays conducted at 1–2 nM PepQ, based on a reproducible lack of detectable enzymatic activity, even with up to eight hours of incubation at 23 °C. PepQ enzymatic activity is, however, detectable when the native dimer is diluted to 1–2 nM (Supplementary Fig. 10). Fluorescence correlation spectroscopy (FCS) data were collected by placing 10 μl of the quenched reaction mixtures onto BSA-blocked coverslips mounted on a custom-built, inverted confocal microscope and covered with a humidified chamber to prevent evaporation.

Autocorrelation curves were collected for 2 min using 50 μW continuous input power from a 561 nm diode-pumped solid state laser. Autocorrelation curves were normalized in mean amplitude between 10^{-6} and 10^{-5} s for display purposes. As standards, the autocorrelation curves of PepQ fully bound to GroEL (obtained by not adding ATP to a folding reaction), as a native dimer (obtained by diluting native PepQ-24TMR in buffer), and as a native monomer (obtained by allowing a GroEL-mediated folding reaction with 1 nM PepQ to continue for an hour) were also determined. Each autocorrelation curve was fit using a multi-component model[64,65] to account for populations of freely diffusing and GroEL-bound PepQ. Curve fitting was conducted using two different approaches. First, the diffusion coefficient of each population was fixed and the fractional population was allowed to vary. Second, the average diffusion coefficient of the entire population was determined. The resulting refolding curves obtained from these two methods were statistically equivalent.

Two-colour single-molecule co-incidence detection. The 24OG and 24TMR PepQ variants were each diluted to 5 μM in acid-urea and allowed to unfold at room temperature. For spontaneous folding, the two solutions were diluted together into acid-urea to a concentration of 50 nM each. Folding was initiated by a 50-fold dilution into TKM buffer containing 0.1 mg ml^{-1} BSA. BSA has no effect on the folding rate of PepQ (Fig. 10b), but was necessary to prevent loss of protein at very-low concentrations to liquid handling equipment. After 10 min at 23 °C, samples were diluted 20-fold into the same buffer and 10 μl samples were immediately placed on a BSA-blocked, optical glass coverslip mounted in a custom holder, fitted on the microscope stage. Samples were covered with a humidifier cap to prevent evaporation. For chaperone-mediated folding, the two 5 μM solutions of denatured, labelled PepQ were diluted together to 50 nM each into TKM containing GroEL and GroES, and folding was initiated by the addition of ATP (2 mM) to a solution containing: 50 nM PepQ-24OG, 50 nM PepQ-24TMR, 200 nM GroEL, 400 nM GroES). After incubated at 23 °C for at least one hour, the reaction was halted by a 50-fold dilution into TKM. This sample was then diluted a further 20-fold into TKM containing 0.1 mg ml^{-1} BSA and immediately assayed. Fluorescence burst data were collected for each sample over a 1 min window using 100 μs sampling bins. Simultaneous excitation was provided from two co-aligned lasers (488 and 561 nm), each providing 200 μW of power at the sample.

To quantify the formation of native PepQ dimers resulting from productive GroEL folding, as well as the formation of PepQ aggregates during spontaneous folding, we developed a cross-correlation statistic (plotted in Fig. 3) that evaluates the percent photon arrival time overlap between two time streams. To begin, each time stream was normalized so the maximum spike intensity amplitude had a value of one. A threshold filter was then applied (5*r.m.s.) to both colour channels to isolate spike activity and remove low-level detector noise. The filtered time streams were used to create a binary mask of spike events. On the basis of a particle transit time through the excitation volume of about 1 ms, both binary time streams were re-binned in 1 ms time bins. The cross-correlation versus time lag between two time streams T_1 and T_2, each with a total of N_1 and N_2 non-zero time bins, was then generated for the + EL/ES/ATP and spontaneous activity data:

$$CC(lag) = \sum \frac{T_1(t - lag) * T_2(t)}{\sqrt{N_1}\sqrt{N_2}} \quad (1)$$

With this normalization, the autocorrelation of any time stream had a value of 1, while the minimum cross-correlation value was bounded at zero. Due to the non-zero probability of photons randomly arriving at two detectors at the same time, the minimal cross-correlation value was not zero.

To assess our cross-correlation measure we used the photon arrival data from either + EL/ES/ATP or spontaneous PepQ activity to generate an expected baseline activity (that is, zero significant co-incidence). Each time stream was compared to a 5 s cyclically shifted version of itself to examine the correlation between two nominally uncorrelated time streams of identical photon rate and noise (denoted as No Overlap, Fig. 3c,d). In both of these baseline cases, there is approximately a 1% cross-correlation independent of lag time or detector channel (red line). We then calibrated the cross-correlation statistic for several data streams of a known and fixed amount of similarity. To do this, we replaced a segment of a time stream T1 with an equal length segment of a time stream T2 at a random location. The ratio of the segment length to total length then corresponded to the percent overlap. The original data stream T2 and the altered data stream T1 then represented two data streams of known overlap and whose cross-correlation could be used for comparison. This process was repeated 20 times for each percent overlap. The cross-correlation results were averaged and the uncertainty in the mean for each lag was monitored. The resulting family of cross-correlation curves (Fig. 3a,b) was then used to assess the level of overlap between two-colour channels that had not been shifted in time.

We then tested the null hypothesis that the pairwise differences between the cross-correlations values of the spontaneous data and various possible overlap simulations had a mean equal to zero. The resulting *P* values from this family of *t*-tests indicated the most likely zero null hypothesis occurred for an overlap of 0.75% (Fig. 3b, inset).

Triplet state conversion of fluorescent PepQ variants. Native, TMR- or Oregon Green-labelled PepQ dimer (24TMR and 24OG, respectively) was diluted to 1 nM (dimer) in TKM buffer with supplemental BSA (0.1 mg ml^{-1}). An amount of 10 μl

of sample was pipetted onto a BSA-blocked coverslip, mounted as described for Fig. 2, using a 561 nm laser for 24TMR and a 488 nm laser for 24OG. For each sample, data were collected for 2 min, with 500 µs collection bins, at each laser power level. The entire, non-normalized data set for each dye (Supplementary Fig. 2) was fit globally using IgorPro (Wavemetrics) to an autocorrelation function that included a correction factor for the effect of a triplet state population[66,67].

Protease protection. The protease sensitivity of non-native PepQ bound to a GroEL ring was conducted as described previously for the substrate protein RuBisCO[23,50]. Briefly, denatured PepQ (100 nM) labelled at position 24 with fluorescein (PepQ-24F) was bound to asymmetric GroEL-GroES-ADP bullets (120 nM)[23]. Chymotrypsin (0.3 mg ml^{-1}) was added, and time points were taken, with the reaction stopped by addition of phenylmethylsulfonyl fluoride (PMSF) (1 mM). Samples were run on 10% SDS–PAGE and imaged using a Typhoon Trio (GE Healthcare).

GroEL ATPase activity. The ATPase activity of GroEL was assayed using an NADH-coupled reaction. In brief, the GroEL (200 nM) ATPase cycle was monitored in the presence of ATP (2 mM) and GroES (400 nM) using 10 U ml^{-1} pyruvate kinase, 10 U ml^{-1} lactate dehydrogenase, 1 mM phosphoenolpyruvate and 0.2 mM NADH. This system regenerates ATP, maintaining it at a constant concentration, and produces a decrease in the absorbance at 340 nm as NADH is consumed[22,39,50,68,69]. The rate of spontaneous ATP hydrolysis under each condition without GroEL was also determined to control for effects on the coupling system.

Cryo-electron microscopy. PepQ was denatured in 8 M urea, 25 mM glycine phosphate, at pH 2 and incubated for 30 min at room temperature. Denatured PepQ (50 µM), in droplets of 4.6 µl (2.3 µM per addition) was titrated into solutions of either GroEL or Δ526 (8 µM tetradecamers, 100 µl) in TKM buffer, followed by rapid, repeated mixing and then incubated at room temperature for 5 min. The final concentration of PepQ was 7 µM. 3 µl of this PepQ/GroEL mixture was applied to a C-Flat 1.2/1.3 400 mesh holey carbon grid at 20 °C with 100% relative humidity and vitrified using a Vitrobot (Mark III, FEI company, Netherlands). The thin-ice areas that showed clear and mono-dispersed particles were imaged under an FEI Tecnai F20 electron microscope with a field emission gun (FEI company, Netherlands) operated at 200 KV. Data were collected on a Gatan K2 Summit direct detection camera (Gatan, Pleasanton CA) in electron counting mode[70] at a nominal magnification of × 19,000, yielding a pixel size of 1.85 Å. The dose rate was 10 e$^-$ pixel^{-2} s^{-1} at the camera. A 33-frame movie stack was recorded for each micrograph, for a total exposure time of 6.6 s. The total dose onto the specimen was 19 e$^-$ Å^{-2}.

Image processing and map visualization. For the wild-type GroEL-PepQ complex, 1,450 micrographs were collected and aligned iteratively and filtered based on electron dose using Unblur[71]. Sum images (1,109) showing strong power spectra were selected for particle picking in EMAN2 (ref. 72), yielding 217,317 particles with a box size of 160 × 160 pixels^2. These particles were processed in Relion[73] with a downscaling factor of 2, respectively. Two runs of reference-free 2D classification were performed to produce a 'cleaner' data set containing 170,639 particles, which were separated into four classes in 3D classification with no symmetry applied. Particles belonging to apo GroEL and PepQ-bound GroEL were used for final asymmetric 3D refinement, respectively (Supplementary Fig. 5). The processing procedure used for the Δ526-PepQ complex was the same as used for the wild type GroEL complex. Briefly, 847 movie stacks were collected and 703 aligned images were picked, yielding 224,696 particles. 117,040 clean particles were screened after two runs of reference-free 2D classification. 3D classification and asymmetric refinement was performed similar as wild-type GroEL-PepQ complex (Supplementary Fig. 6). The final resolutions of the 3D density maps for both wild-type GroEL and Δ526 were 7.9 Å for the apo states and 8.3 Å for the PepQ-bound states (Supplementary Fig. 7A), assessed with the gold-standard criterion at 0.143 Fourier shell correlation[74]. Local resolutions were estimated using Blocres[75]. The unwrapping of the maps was done with 'e2unwrap3d.py' in EMAN2. Visualization and fitting of atomic models into the cryo-EM density maps, were done in UCSF Chimera[76].

Circular dichroism spectroscopy. PepQ was diluted >100-fold into 25 mM sodium phosphate, pH 7.2, 100 µM $MnCl_2$ (native) or 25 mM sodium phosphate, pH 2.1, 8 M Urea (denatured) to a concentration of 0.1 mg ml^{-1}. Following an equilibration at 23 °C for 15 min, samples were loaded into a 1 cm path length cuvette and the circular dichroism (CD) of the sample measured in the far ultra-violet region using an Aviv 202:CD spectrometer (Aviv Biomedical). The sample temperature was equilibrated in the spectrometer to 25 °C before the initiation of measurements and was maintained at this temperature throughout. The CD signal at each wavelength was averaged for 30 s, using 1 nm wavelength steps. Scans of each sample buffer were used as blanks.

Encapsulation of PepQ by GroEL. PepQ-24F (100 nM) was denatured in 8 M urea, 25 mM glycine phosphate, pH 2 and bound to wild-type or Δ526 GroEL-GroES-ADP bullets (200 nM) or full-length or Δ526 single-ring GroEL (300 nM) supplemented with GroES (600 nM). A single turnover was initiated by the addition of ATP (2 mM) followed by quenching with hexokinase and glucose after 10 s. Unencapsulated PepQ was digested with Proteinase K (0.5 µg ml^{-1}) for 10 min before quenching with PMSF (1 mM). Samples were run on SDS–PAGE and scanned for fluorescein fluorescence using a Typhoon Trio (GE Healthcare). Band intensity was measured with ImageJ.

Data availability. The cryo-EM density maps are deposited to EMDataBank (http://www.emdatabank.org/) with accession id EMD-8316 (wild type GroEL with PepQ bound), EMD-8317 (wild type GroEL), EMD-8318(C-terminal truncated GroEL, Δ526, with PepQ bound) and EMD-8319(C-terminal truncated GroEL, Δ526). All other data are available from the corresponding authors upon reasonable request.

References

1. Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M. & Hartl, F. U. Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **82,** 323–355 (2013).
2. Horwich, A. L. & Fenton, W. A. Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. *Q. Rev. Biophys.* **42,** 83–116 (2009).
3. Lin, Z. & Rye, H. GroEL-mediated protein folding: making the impossible, possible. *Crit. Rev. Biochem. Mol. Bio* **41,** 211–239 (2006).
4. Yébenes, H., Mesa, P., Muñoz, I. G., Montoya, G. & Valpuesta, J. M. Chaperonins: two rings for folding. *Trends Biochem. Sci.* **36,** 424–432 (2011).
5. Hayer-Hartl, M., Bracher, A. & Hartl, F. U. The GroEL-GroES chaperonin machine: a nano-cage for protein folding. *Trends Biochem. Sci.* **41,** 62–76 (2016).
6. Braig, K. *et al.* The crystal structure of the bacterial chaperonin GroEL at 2.8A. *Nature* **371,** 578–586 (1994).
7. Fenton, W., Kashi, Y., Furtak, K. & Horwich, A. Residues in chaperonin GroEL required for polypeptide binding and release [see comments]. *Nature* **371,** 614–619 (1994).
8. Chen, L. & Sigler, P. The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. *Cell* **99,** 757–768 (1999).
9. Weissman, J., Rye, H., Fenton, W., Beechem, J. & Horwich, A. Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. *Cell* **84,** 481–490 (1996).
10. Mayhew, M. *et al.* Protein folding in the central cavity of the GroEL-GroES chaperonin complex. *Nature* **379,** 420–426 (1996).
11. Weissman, J. *et al.* Mechanism of GroEL action: productive release of polypeptide from a sequestered position under GroES. *Cell* **83,** 577–587 (1995).
12. Liu, C. *et al.* Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* **463,** 197–202 (2010).
13. Todd, M. J., Viitanen, P. V. & Lorimer, G. H. Dynamics of the chaperonin ATPase cycle: implications for facilitated protein folding. *Science* **265,** 659–666 (1994).
14. Weissman, J., Kashi, Y., Fenton, W. & Horwich, A. GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. *Cell* **78,** 693–702 (1994).
15. Jewett, A. I. & Shea, J.-E. Reconciling theories of chaperonin accelerated folding with experimental evidence. *Cell. Mol. Life Sci.* **67,** 255–276 (2010).
16. Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475,** 324–332 (2011).
17. Todd, M., Lorimer, G. & Thirumalai, D. Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism. *Proc. Natl Acad. Sci. USA* **93,** 4030–4035 (1996).
18. Betancourt, M. & Thirumalai, D. Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity. *J. Mol. Biol.* **287,** 627–644 (1999).
19. Lin, Z. & Rye, H. S. Expansion and compression of a protein folding intermediate by GroEL. *Mol. Cell* **16,** 23–34 (2004).
20. Brinker, A. *et al.* Dual function of protein confinement in chaperonin-assisted protein folding. *Cell* **107,** 223–233 (2001).
21. Chakraborty, K. *et al.* Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. *Cell* **142,** 112–122 (2010).
22. Lin, Z., Puchalla, J., Shoup, D. & Rye, H. S. Repetitive protein unfolding by the trans ring of the GroEL-GroES chaperonin complex stimulates folding. *J. Biol. Chem.* **288,** 30944–30955 (2013).
23. Lin, Z., Madan, D. & Rye, H. S. GroEL stimulates protein folding through forced unfolding. *Nat. Struct. Mol. Biol.* **15,** 303–311 (2008).
24. Sharma, S. *et al.* Monitoring protein conformation along the pathway of chaperonin-assisted folding. *Cell* **133,** 142–153 (2008).
25. Georgescauld, F. *et al.* GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding. *Cell* **157,** 922–934 (2014).

26. Ambrose, A. J., Fenton, W., Mason, D. J., Chapman, E. & Horwich, A. L. Unfolded DapA forms aggregates when diluted into free solution, confounding comparison with folding by the GroEL/GroES chaperonin system. *FEBS Lett.* **589,** 497–499 (2015).
27. Weaver, J., Watts, T., Li, P. & Rye, H. S. Structural basis of substrate selectivity of E. coli prolidase. *PLoS ONE* **9,** e111531 (2014).
28. Park, M.-S. *et al.* Catalytic properties of the PepQ prolidase from *Escherichia coli*. *Arch. Biochem. Biophys.* **429,** 224–230 (2004).
29. Bazan, J. F., Weaver, L. H., Roderick, S. L., Huber, R. & Matthews, B. W. Sequence and structure comparison suggest that methionine aminopeptidase, prolidase, aminopeptidase P, and creatinase share a common fold. *Proc. Natl Acad. Sci. USA* **91,** 2473–2477 (1994).
30. Lowther, W. T. & Matthews, B. W. Structure and function of the methionine aminopeptidases. *Biochim. Biophys. Acta* **1477,** 157–167 (2000).
31. Kerner, M. *et al.* Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122,** 209–220 (2005).
32. Fujiwara, K., Ishihama, Y., Nakahigashi, K., Soga, T. & Taguchi, H. A systematic survey of *in vivo* obligate chaperonin-dependent substrates. *EMBO J.* **29,** 1552–1564 (2010).
33. Grallert, H., Rutkat, K. & Buchner, J. Limits of protein folding inside GroE complexes. *J. Biol. Chem.* **275,** 20424–20430 (2000).
34. Tang, Y. *et al.* Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. *Cell* **125,** 903–914 (2006).
35. Goloubinoff, P., Christeller, J. T., Gatenby, A. A. & Lorimer, G. H. Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature* **342,** 884–889 (1989).
36. van der Vies, S. M., Viitanen, P. V., Gatenby, A. A., Lorimer, G. H. & Jaenicke, R. Conformational states of ribulosebisphosphate carboxylase and their interaction with chaperonin 60. *Biochemistry* **31,** 3635–3644 (1992).
37. Rye, H. *et al.* GroEL-GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings. *Cell* **97,** 325–338 (1999).
38. Burston, S. G., Ranson, N. A. & Clarke, A. R. The origins and consequences of asymmetry in the chaperonin reaction cycle. *J. Mol. Biol.* **249,** 138–152 (1995).
39. Weaver, J. & Rye, H. S. The C-terminal tails of the bacterial chaperonin GroEL stimulate protein folding by directly altering the conformation of a substrate protein. *J. Biol. Chem.* **289,** 23219–23232 (2014).
40. Chen, D.-H. *et al.* Visualizing GroEL/ES in the act of encapsulating a folding protein. *Cell* **153,** 1354–1365 (2013).
41. Tang, Y., Chang, H., Chakraborty, K., Hartl, F. & Hayer-Hartl, M. Essential role of the chaperonin folding compartment *in vivo*. *EMBO J.* **27,** 1458–1468 (2008).
42. McLennan, N. F., Girshovich, A. S., Lissin, N. M., Charters, Y. & Masters, M. The strongly conserved carboxyl-terminus glycine-methionine motif of the *Escherichia coli* GroEL chaperonin is dispensable. *Mol. Microbiol.* **7,** 49–58 (1993).
43. Burnett, B. P., Horwich, A. L. & Low, K. B. A carboxy-terminal deletion impairs the assembly of GroEL and confers a pleiotropic phenotype in *Escherichia coli* K-12. *J. Bacteriol.* **176,** 6980–6985 (1994).
44. Machida, K., Kono-Okada, A., Hongo, K., Mizobata, T. & Kawata, Y. Hydrophilic residues 526KNDAAD531 in the flexible C-terminal region of the chaperonin GroEL are critical for substrate protein folding within the central cavity. *J. Biol. Chem.* **283,** 6886–6896 (2008).
45. Farr, G., Fenton, W. & Horwich, A. Perturbed ATPase activity and not 'close confinement' of substrate in the cis cavity affects rates of folding by tail-multiplied GroEL. *Proc. Natl Acad. Sci. USA* **104,** 5342–5347 (2007).
46. Brocchieri, L. & Karlin, S. Conservation among HSP60 sequences in relation to structure, function, and evolution. *Protein Sci.* **9,** 476–486 (2000).
47. Lund, P. A. Multiple chaperonins in bacteria--why so many? *FEMS Microbiol. Rev.* **33,** 785–800 (2009).
48. Elad, N. *et al.* Topologies of a substrate protein bound to the chaperonin GroEL. *Mol. Cell* **26,** 415–426 (2007).
49. Gruber, R. & Horovitz, A. Allosteric mechanisms in chaperonin machines. *Chem. Rev.* **116,** 6588–6606 (2016).
50. Madan, D., Lin, Z. & Rye, H. S. Triggering protein folding within the GroEL-GroES complex. *J. Biol. Chem.* **283,** 32003–32013 (2008).
51. Grason, J. P., Gresham, J. S. & Lorimer, G. H. Setting the chaperonin timer: a two-stroke, two-speed, protein machine. *Proc. Natl Acad. Sci. USA* **105,** 17339–17344 (2008).
52. Falke, S., Tama, F., Brooks, III C. L., Gogol, E. P. & Fisher, M. T. The 13 Å Structure of a chaperonin GroEL–protein substrate complex by cryo-electron microscopy. *J. Mol. Biol.* **348,** 219–230 (2005).
53. Apetri, A. C. & Horwich, A. L. Chaperonin chamber accelerates protein folding through passive action of preventing aggregation. *Proc. Natl Acad. Sci. USA* **105,** 17351–17355 (2008).
54. Horwich, A. L., Apetri, A. C. & Fenton, W. A. The GroEL/GroES cis cavity as a passive anti-aggregation device. *FEBS Lett.* **583,** 2654–2662 (2009).
55. Libich, D. S., Tugarinov, V. & Clore, G. M. Intrinsic unfoldase/foldase activity of the chaperonin GroEL directly demonstrated using multinuclear relaxation-based NMR. *Proc. Natl Acad. Sci. USA* **112,** 8817–8823 (2015).
56. Voziyan, P. A. & Fisher, M. T. Chaperonin-assisted folding of glutamine synthetase under nonpermissive conditions: off-pathway aggregation propensity does not determine the co-chaperonin requirement. *Protein. Sci.* **9,** 2405–2412 (2000).
57. Xu, Z., Horwich, A. & Sigler, P. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex [see comments]. *Nature* **388,** 741–750 (1997).
58. Roseman, A. M., Chen, S., White, H., Braig, K. & Saibil, H. R. The chaperonin ATPase cycle: mechanism of allosteric switching and movements of substrate-binding domains in GroEL. *Cell* **87,** 241–251 (1996).
59. Chen, S. *et al.* Location of a folding protein and shape changes in GroEL-GroES complexes imaged by cryo-electron microscopy. *Nature* **371,** 261–264 (1994).
60. Motojima, F., Chaudhry, C., Fenton, W. A., Farr, G. W. & Horwich, A. L. Substrate polypeptide presents a load on the apical domains of the chaperonin GroEL. *Proc. Natl Acad. Sci. USA* **101,** 15005–15012 (2004).
61. Dalton, K. M., Frydman, J. & Pande, V. S. The dynamic conformational cycle of the group I chaperonin C-termini revealed via molecular dynamics simulation. *PLoS ONE* **10,** e0117724 (2015).
62. Braman, J., Papworth, C. & Greener, A. Site-directed mutagenesis using double-stranded plasmid DNA templates. *Methods Mol. Biol.* **57,** 31–44 (1996).
63. Rye, H. Application of fluorescence resonance energy transfer to the GroEL-GroES chaperonin reaction. *Methods* **24,** 278–288 (2001).
64. Sengupta, P., Garai, K., Balaji, J., Periasamy, N. & Maiti, S. Measuring size distribution in highly heterogeneous systems with fluorescence correlation spectroscopy. *Biophys. J.* **84,** 1977–1984 (2003).
65. Pal, N., Dev Verma, S., Singh, M. K. & Sen, S. Fluorescence correlation spectroscopy: an efficient tool for measuring size, size-distribution and polydispersity of microemulsion droplets in solution. *Anal. Chem.* **83,** 7736–7744 (2011).
66. Widengren, J. & Rigler, R. Mechanisms of photobleaching investigated by fluorescence correlation spectroscopy. *Bioimaging* **4,** 149–157 (1996).
67. Widengren, J., Rigler, R. & Mets, U. Triplet-state monitoring by fluorescence correlation spectroscopy. *J. Fluoresc.* **4,** 255–258 (1994).
68. Poso, D., Clarke, A. & Burston, S. A kinetic analysis of the nucleotide-induced allosteric transitions in a single-ring mutant of GroEL. *J. Mol. Biol.* **338,** 969–977 (2004).
69. Kreuzer, K. N. & Jongeneel, C. V. *Escherichia coli* phage T4 topoisomerase. *Methods Enzymol.* **100,** 144–160 (1983).
70. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10,** 584–590 (2013).
71. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4,** e06980 (2015).
72. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157,** 38–46 (2007).
73. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180,** 519–530 (2012).
74. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9,** 853–854 (2012).
75. Heymann, J. B., Cardone, G., Winkler, D. C. & Steven, A. C. Computational resources for cryo-electron tomography in Bsoft. *J. Struct. Biol.* **161,** 232–242 (2008).
76. Pettersen, E. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
77. Yamamoto, D. & Ando, T. Chaperonin GroEL-GroES functions as both alternating and non-alternating engines. *J. Mol. Biol.* **428,** 3090–3101 (2016).
78. Fei, X., Ye, X., LaRonde, N. A. & Lorimer, G. H. Formation and structures of GroEL:GroES2 chaperonin footballs, the protein-folding functional form. *Proc. Natl Acad. Sci. USA* **111,** 12775–12780 (2014).
79. Haldar, S. *et al.* Chaperonin-assisted protein folding: relative population of asymmetric and symmetric GroEL:GroES complexes. *J. Mol. Biol.* **427,** 2244–2255 (2015).
80. Grason, J. P., Gresham, J. S., Widjaja, L., Wehri, S. C. & Lorimer, G. H. Setting the chaperonin timer: the effects of K+ and substrate protein on ATP hydrolysis. *Proc. Natl Acad. Sci. USA* **105,** 17334–17338 (2008).

Acknowledgements

This research has been supported by the National Institutes of Health grant (H.R. GM065421). We thank the Microscopy and Imaging Center at Texas A&M University for providing instrumentation for cryo-EM data collection. We acknowledge the Texas A&M High Performance Research Computing Center for providing the computational resources for the data processing. We also thank Dr Chavela Carr for comments and editorial assistance in preparing the manuscript. J.Z. is supported by startup funding from the Department of Biochemistry and Biophysics at Texas A&M

University and Center for Phage Technology jointly sponsored by Texas AgriLife and Texas A&M University. J.Z. is also supported by Welch Foundation grant A-1863.

Author contributions

Conceptualization: H.S.R., J.Z. and J.W. Methodology: H.S.R., J.Z., J.W., A.R. and J.P. Formal analysis: J.P., J.Z and M.J. Investigation: J.W., A.R. and M.J. Writing—original draft: H.S.R. Writing—review and editing: H.S.R., J.Z., J.P., J.W., A.R. and M.J. Visualization: H.S.R., J.Z., J.W., J.P. and M.J. Supervision: H.S.R and J.Z. Project management: H.S.R. Funding acquisition: H.S.R and J.Z.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Weaver, J. *et al.* GroEL actively stimulates folding of the endogenous substrate protein PepQ. *Nat. Commun.* **8,** 15934 doi: 10.1038/ncomms15934 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.