Structure of the cell-binding component of the *Clostridium difficile* binary toxin reveals a di-heptamer macromolecular assembly

Xingjian Xu,a,b, Raquel Godoy-Ruiz,c,d,e, Kaylin A. Adipietro,c,e Christopher Peralta,a, Danya Ben-Hai,a, Kristen M. Varney,c,d,e, Mary E. Cook,c,e, Braden M. Roth,b, Paul T. Wilder,c,d,e, Thomas Cleveland,b, Alexander Grishaev,f, Heather M. Neug, Sarah L. J. Michelg, Wenbo Yu,e,g, Dorothy Beckett,h, Richard R. Rustandi, i Catherine Lancaster, i John W. Loughney, e Adam Kristopen, i, Ianny Christianti,i, Jessica W. Olson, i Alexander D. MacKerell,e, S.L.J.M., W.Y., D.B., A.D.M., A.d.G., E.P., and D.J.W. analyzed data; and A.D.M., A.d.G., E.P., and D.J.W. performed research; W.Y., R.R.R., D.B.-H., K.M.V., M.E.C., B.M.R., P.T.W., T.C., A.G., H.M.N., S.L.J.M., W.Y., D.B., R.R.R., C.L., J.W.L., A.K., S.C., J.W.O., A.D.M., A.D.G., E.P., and D.J.W. performed research; W.Y., R.R.R., C.L., J.W.L., A.K., S.C., J.W.O., A.D.M., and E.P. contributed new reagents/analytic tools; X.X., R.R.R., A.G., H.M.N., S.L.J.M., W.Y., D.B., R.R.R., C.L., J.W.L., A.K., S.C., J.W.O., A.D.M., and E.P. performed research; A.D.M., A.d.G., E.P., and D.J.W. wrote the paper.

Edited by Michael F. Summers, University of Maryland, Baltimore, MD, and approved December 10, 2019 (received for review November 7, 2019)

Targeting *Clostridium difficile* infection is challenging because treatment options are limited, and high recurrence rates are common. One reason for this is that hypervirulent *C. difficile* strains often have a binary toxin termed the *C. difficile* toxin, in addition to the enterotoxins TsdA and TsdB. The *C. difficile* toxin has an enzymatic component, termed CDTa, and a pore-forming or delivery subunit termed CDTb. CDTb was characterized here using a combination of single-particle cryoelectron microscopy, X-ray crystallography, NMR, and other biophysical methods. In the absence of CDTa, 2 di-heptamer structures for activated CDTb (1.0 Ma) were solved at atomic resolution, including a symmetric (AsymCDTb; 3.14 Å) and an asymmetric form (1oCANC1b; 2.84 Å). Roles played by 2 receptor-binding domains of activated CDTb were of particular interest since the receptor-binding domain 1 lacks sequence homology to any other known toxin, and the receptor-binding domain 2 is completely absent in other well-studied heptameric toxins (i.e., anthrax). For AsymCDTb, a Ca2+ binding site was discovered in the first receptor-binding domain that is important for its stability, and the second receptor-binding domain was found to be critical for host cell toxicity and the di-heptamer fold for both forms of activated CDTb. Together, these studies represent a starting point for developing structure-based drug-design strategies to target the most severe strains of *C. difficile*.

Antibiotic and antitoxin combination therapy is often an effective clinical approach for toxin-producing infections (10), so this strategy is under development for treating *C. difficile* infection. While therapeutic options are becoming available to

**Significance**

There is a burden from *Clostridium difficile* infection throughout the world, and the Centers for Disease Control reports more than 500,000 cases annually in the United States, resulting in an estimated 15,000 deaths. In addition to the large clostridial toxins, TcdA/TcdB, a third *C. difficile* binary toxin is associated with the most serious outbreaks of drug-resistant *C. difficile* infection in the 21st century. Here, structural biology and biophysical approaches were used to characterize the cell binding component of the *C. difficile* binary toxin, termed CDTb. Surprisingly, 2 structures were solved from a single sample that help explain the molecular underpinnings of *C. difficile* toxicity. These structures will also be important for targeting this human pathogen via structure-based therapeutic design methods.

*Clostridium difficile* | cryo-EM | NMR | X-ray crystallography | structural biology

Symbiotic microbiota in the gut typically prevent *Clostridium difficile* colonization in healthy individuals, but as protective bacteria are reduced by common antibiotic treatments, cancer therapy, and by other means, then *C. difficile* infection becomes a much higher health risk (1, 2). Upon diagnosis, it is critical to cease delivery of problematic antibiotics, particularly those prone to select for hypervirulent strains (i.e., fluoroquinolones, clindamycin, cephalosporins) (3, 4), and then clear the infection with a limited choice of antibiotics that can sometimes provide efficacy, including metronidazole, vancomycin, and fidaxomicin (1, 5). However, continued resistance to antibiotics and overwhelming levels of toxin production by the *C. difficile* bacteria can severely limit such a clinical approach. Other options for patients having severe *C. difficile* infection are colonoscopy or experimental procedures, such as a fecal microbiota transplant, but these treatment options can have severe drawbacks (1, 6). Consequently, novel therapies are needed, particularly for recurrent *C. difficile* infection and for cases associated with hypervirulent strains (i.e., BI, NAP1, 027, 078, and others) (1, 5, 7–9).
target the large clostridial toxins, TcdA/TcdB (11), there is nothing approved by the Food and Drug Administration to target the C. difficile toxin (CDT) or the “binary toxin” (12). Other evidence demonstrating an urgency to develop antitoxins targeting the binary toxin include: 1) Patients with binary toxin-containing strains of C. difficile infection show heightened disease severity and reoccurrence (13–16); 2) strains of having only the C. difficile binary toxin and not TcdA/TcdB (A’ B’ CDT’) retain virulence and present as C. difficile infection in the clinic (16, 17), and 3) an immunological response in hamsters to a vaccine targeting TcdA/TcdB and the binary toxin showed much higher efficacy toward challenges from a hypervirulent strain of C. difficile infection (i.e., NAP1) than a vaccine derived only from TcdA/TcdB antigens (12, 18). Therefore, to address this unmet medical need, studies of the structure, function, and inhibition of the C. difficile binary toxin are paramount to identifying its vulnerabilities and for developing novel treatments to improve patient outcomes for the most severe cases of C. difficile infection.

The CDT is a binary toxin that has an enzymatic subunit, CDTa (47.4 kDa), with ribosyltransferase activity and a pore-forming delivery subunit, termed CDTb (99 kDa) (15, 19–23). Prior to cellular entry via endosomes (24–27), the CDT associates with host cell receptors, such as the lipopolysaccharide-stimulated lipoprotein receptor and CD8 (28–31). Based on studies with other binary toxins, it was suggested that the low pH in endosomes triggers CDTa translocation into the cytoplasm, via the cell-binding and pore-forming entity, CDTb, but a detailed molecular mechanism for this process remains unknown (32–41). Once the CDTa enzyme is delivered into the host cell cytoplasm, ADP ribosylation of G-actin occurs catalytically at Arg-177 (42). ADP ribosylated G-actin then leads to F-actin filament dissociation (43), destruction of the cytoskeleton, increased microtubule protrusions, accelerated bacterial adhesion, and a “death spiral” for host cells (44–46). In this study, a combination of biophysical and structural biology methods was used to define the molecular structure of activated CDTb. The roles played by the 2 receptor-binding domains of CDTb were of particular interest in this study. Receptor-binding domain 1 (RBD1) lacks sequence homology to any other known toxin and was found to have a Ca2+-binding site. The second RBD is at the C terminus of CDTb, and it is not present in other members of this toxin family. Importantly, RBD2 was shown to be critical for establishing the di-heptamer macromolecular assembly in activated CDTb that is necessary for host cell toxicity. Together, these and other regions of activated CDTb can now be considered in future mechanistic studies as well as in structure-based drug design strategies.

Results and Discussion

Structural and Biophysical Characterization of Active CDTb. For studies of activated CDTb, inactive pro-CDTb (residues 1 to 876) was overexpressed first in baculovirus-infected insect cells and purified to homogeneity. Active CDTb (residues 212 to 876) was generated via limited proteolysis of pro-CDTb with chymotrypsin to remove the signaling peptide (residues 1 to 43) and the activation domain (residues 43 to 211) with hydrolysis confirmed to be between M211 and S212 by mass spectroscopy, as previously described (12). Activated CDTb protein was purified to homogeneity (>99%) and shown to be fully toxic in Vero cell-killing assays using catalytic amounts of activated CDTa (CDT a\textsubscript{TC50} = 110 ± 10 pM) (Fig. 1) and an optimal CDTa to CDTb ratio of 1.7, as was previously described (47).

Sizing studies of active CDTb were completed using sedimentation velocity analytical ultracentrifugation (AUC) and size-exclusion chromatography/multangle light scattering (SEC-MALS) to determine its subunit stoichiometry. Surprisingly, rather being heptamer, as described for other cell-binding components of binary toxins (41), both methods showed that monomeric CDTb (75 kDa) was the major species (95 ± 2%) and a novel 14-subunit oligomer (1.0 MDa) was detected at lower levels (<4 to 6%) (Fig. 2A and B). Interestingly, the addition of CDTa to activated CDTb shifted equilibrium toward larger species (>25%) (Fig. 2A) together with increased size of the complex (1.1 MDa), as expected from predicted 1:7 CDTa:CDTb stoichiometry. There was no evidence for heptameromic CDTb after its activation or upon the addition of CDTa (<0.1%). In the absence of CDTa, the presence of the 14-subunit CDTb oligomer was validated using small-angle X-ray scattering (SAXS), and consistent with the SEC-MALS and AUC data, the SAXS data indicated that 14mer has radius of gyration of 86 ± 2 Å and a molecular weight of 1.0 ± 0.2 MDa (Fig. 2C). The interatomic distance probability distribution calculated from the SAXS scattering profile indicated that activated CDTb had a maximum particle dimension of 270 Å, and modeling these data with 2 dumbbell-like shapes of the di-heptamer markedly improved the quality of fit.

Importantly, single-particle cryoelectron microscopy (cryo-EM) studies were completed for activated CDTb (SI Appendix, Figs. S1 and S2), in the absence of CDTa, and demonstrated unambiguously that the 14mer oligomerization state was the only higher molecular weight state observed, but interestingly, it had 2 unique structures including a symmetric (\textsuperscript{sym}CDTb) and an asymmetric (\textsuperscript{asym}CDTb) dimer of heptamers, which were solved at resolutions of 3.14 Å and 2.84 Å, respectively (Figs. 3–7). Similarly, crystals were obtained from the same active CDTb preparations and the availability of cryo-EM models was essential to solving its structure by molecular replacement. The results of the X-ray studies further confirmed the dimer of heptamer stoichiometry for CDTb; however, only the \textsuperscript{asym}CDTb was observed when the X-ray diffraction data were analyzed at 3.70-Å resolution (SI Appendix, Fig. S3). These structures of activated CDTb will be important for continued delineation of the toxin’s mechanism of action as well as for future drug development efforts targeting the CDT.

\textsuperscript{sym}CDTb and \textsuperscript{asym}CDTb at Atomic Resolution. The X-ray and cryo-EM structures of the cell binding and delivery component of the binary toxin, CDTb, were examined in detail. Single-particle cryo-EM studies of active CDTb revealed 2 unique structures, including...
a symmetric ($^{Sym}$CDTb) and an asymmetric ($^{Asym}$CDTb) form (Fig. 3), and the $^{Asym}$CDTb form was confirmed via X-ray crystallography (SI Appendix, Fig. S3).

The Global Folds of the 2 Di-heptamer Active CDTb Structures. The heptamer units in the di-heptamers of CDTb assume 2 distinct forms. As shown (Figs. 3–5), an extended β-barrel resides in 1 of the heptamer units that resembles the low pH membrane inserted structure of the protective antigen cell-binding component of the anthrax toxin (48, 49), while the other lacks this structural motif and is more similar to the soluble form of the anthrax toxin. Nonetheless, while there are some similarities, the structures of both CDTb di-heptamers differs significantly from the heptameric assembly characteristic of the pore-forming component of the anthrax-protective antigen (Fig. 5) (41, 48, 50–52). Specifically, for $^{Sym}$CDTb, both heptamers of the di-heptamer are in a non-β-barrel form. The non-β-barrel/non-β-barrel assembly of the 2 heptamers for $^{Sym}$CDTb is driven by a central donut-like structure formed by 14 copies of the 14-kDa C-terminal domain of CDTb, termed the RBD2, which is absent in the anthrax protective antigen (Fig. 5). Whereas, $^{Asym}$CDTb comprises a mixed non-β-barrel/β-barrel di-heptamer assembly, but again the 2 heptamer assemblies of this asymmetric form are brought together as a di-heptamer by this unique RBD2-mediated mechanism. In $^{Asym}$CDTb, the 105-Å-long β-barrel structure makes additional non-β-barrel/β-barrel interactions with the RBD2 domains (Fig. 7A) and shield several hydrophobic residues, which likely stabilizes $^{Asym}$CDTb prior to CDTa/receptor binding and insertion into the lipid membrane of host cells.

Specific Domain Structures within $^{Asym}$CDTb and $^{Sym}$CDTb. For more detailed comparisons, delineation of domains of active CDTb are based on homologous domains from heptameric toxins (Fig. 3). These include a heptamerization domain (HD1; residues 212 to 297), the β-barrel domain (βBD; residues 298 to 401), a second heptamerization domain (HD2; residues 402 to 486), a linker region (L1; residues 487 to 513), a third heptamerization domain (HD3; residues 514 to 615), an RBD (RBD1; residues 616 to 744), a second linker (L2; residues 745 to 756), and a second RBD (RBD2; residues 757 to 876) (SI Appendix, Figs. S4–S11). It is important to point out that RBD1 is not homologous to any other binary toxin, and when aligned, no other toxin was found to have sequence homologous to RBD2. On the other hand, as for other heptameric pore-forming toxins, HDs 1, 2, and 3 plus RBD1 comprise a large number of the interdomain interactions within a single heptamer unit in both $^{Sym}$CDTb and $^{Asym}$CDTb (Fig. 3).

The Heptamer Core of $^{Asym}$CDTb and $^{Sym}$CDTb. We refer to the 3 heptamer domains of active CDTb as the heptamer core because these regions of the toxin retain folds similar to that observed for other toxins in this class with their sequences aligning with up to ~20% identity. The first heptamerization domain belongs to the clostridial calcium binding domain family (53). HD1 occurs in both $^{Sym}$CDTb and $^{Asym}$CDTb and features 2 proximal Ca2+ binding sites (Fig. 6A) that are highly conserved in this toxin family (54, 55). The presence of Ca2+ was confirmed here for active CDTb using inductively coupled plasma mass spectrometry. These Ca2+-binding sites do play a structural role in anthrax toxin (54, 56) and extracellular calcium is required for several steps in the intoxication of anthrax and iota toxin in cell-based assays (57, 58), so these results were not too surprising here for active CDTb. The first heptamerization domain is followed by what is termed the βBD, and it is this domain that establishes the ~105-Å-long β-barrel structure that is observed in what is termed here the “β-barrel heptamer unit.” Specifically, 2 strands from 7 subunits of active CDTb are elongated into 70-residue-long double-stranded antiparallel β-sheet that together
Fig. 3. Structures of activated CDTb. (A) Local resolution in structures of \textsuperscript{Asym}CDTb and \textsuperscript{Sym}CDTb conformations. Increased flexibility is observed in outer regions of the core heptamer, most pronounced for the RBD1 domain. (B) Overall structure of the activated CDTb tetradecamer in \textsuperscript{Asym}CDTb and \textsuperscript{Sym}CDTb conformations. Color scheme is shown in domain diagram and both models are on the same scale, demonstrating slight shortening of the \textsuperscript{Sym}CDTb. Domains include a heptamerization domain (HD1; residues 212 to 297), a \(\beta\)BD (residues 298 to 401), a second heptamerization domain (HD2; residues 402 to 486), a linker region (L1; residues 487 to 513), a third heptamerization domain (HD3; residues 514 to 615), an RBD (RBD1; residues 616 to 744), a second linker (L2; residues 745 to 756), and a second RBD (RBD2; residues 757 to 876). The secretion peptide (SP) and the activation domain (AD) are removed via chymotrypsin processing to provide activated CDTb (see also SI Appendix, Figs. S4–S11).
form this striking β-barrel fold. At the tip of the β-barrel, there are several hydrophobic residues that are partially protected from solvent via insertion into a cavity that presumably stabilizes AsymCDTb prior to CDTa binding and insertion into the lipid membrane of host cells (Fig. 7A). While the β-barrel structure observed here for AsymCDTb is reminiscent of the pore-forming component of the protective antigen of the anthrax toxin, it is important to emphasize that in the case of AsymCDTb, it does not require a lipid bilayer or presence of detergents to form. In the “non–β-barrel heptamer,” these same residues have a drastically different structure, as it retains a 4-stranded antiparallel β-sheet that packs against the HDs 2 and 3, and this β-sheet structure is interrupted by a long loop that packs in between the third heptamerization domain and the first RBD.

The second heptamerization domain of active CDTb has 2 antiparallel β-strands followed by a 40-residue-long loop, a short α-helix, and a third β-strand, which completes a 3-stranded antiparallel β-sheet in both the non–β-barrel/β-barrel heptamers (SI Appendix, Fig. S8). However, because the second heptamerization domain packs into the βBD, a rigid body type shift in all 3 β-strands and the short helix of the second heptamer domain are observed, which essentially “clamps down” on 2 strands of βBD to provide the unique packing of the β-barrel-forming heptamer, whereas in the non–β-barrel heptamer, these same residues adopt a more “open” conformation and pack against all 4 strands of the β-sheet of the βBD. Remarkably, this subtle differences in structure for residues in this domain are sufficient to reorient a key φ-gate residue, Phe-455, which is functionally important for transporting CDTa through the CDTb pore (37, 48). Thus, 7 phenylalanine residues in the φ-gate of β-barrel unit form a 3-Å orifice in comparison to the non–β-barrel units in which the pore diameter comprising these same phenylalanine residues is 12.5 Å (Fig. 7 B and C). The final components of the heptamer core comprise a short 3-kDa turn-helix linker domain and the third heptamerization domain. The third heptamer domain contains a 4-strand β-sheet flanked by an extended loop region and 2 α-helices, and like the other 2 heptamerization domains, this third heptamer domain contributes to the large CDTa binding cavity just prior to the φ-gate (SI Appendix, Fig. S9). In the third heptamer domain, there are no significant structural differences between non–β-barrel and β-barrel heptamer units (rmsd of 0.32 Å) for SymCDTb and AsymCDTb, respectively, so how this domain affects binding of CDTa to both forms of activated CDTb is of particular interest.

The Receptor Binding Domains of AsymCDTb and SymCDTb. The first RBD is unique to activated CDTb and has no sequence or structural similarity with any corresponding domains from anthrax toxin or any other binary toxin of known structure. RBD1 is a 10-stranded β-sandwich having a fold most similar to what are termed bacterial carbohydrate-binding modules (SI Appendix, Fig. S10). A number of β-sandwich carbohydrate-binding modules are reported to bind calcium ions (59), as was observed here for the RBD1 in the X-ray crystal structure of AsymCDTb (Fig. 7B). Interestingly, RBD1 is better resolved in the crystal structure since it is stabilized by crystal contacts, whereas evidence for Ca²⁺ occupancy in the same location in cryo-EM density is somewhat obscure due to increased flexibility of these regions in
solution. Second, when the sequence comprising the RBD1 was isolated (residues 616 to 744), it was found to be unfolded as determined by severe line-broadening effects and the lack of chemical-shift dispersion in a $^{15}$N-edited HSQC NMR experiment; however, upon the addition of Ca$^{2+}$, the line-width values narrowed and significant chemical-shift dispersion appeared that is typical of a fully folded protein. Importantly, evaluation of chemical-shift indices in the NMR data illustrate that Ca$^{2+}$-bound form of the RBD1 folds into a secondary structure that is in full agreement with that observed using cryo-EM and X-ray crystallography of the full-length construct (Fig. 6).

The second RBD at the toxin’s C terminus is connected to RBD1 by a 12-residue linker (residues 745 to 756). Little if any change in its fold is observed when this domain is compared among all of the heptamer units (rmsd of 0.35 Å) or to a crystal structure of the RBD2 determined here in isolation (SI Appendix, Fig. S11). When the $^{\text{Sym}}$CDTb and $^{\text{Asym}}$CDTb structures are compared, however, the location of the second RBD is very different. Specifically, in the β-barrel heptamer of $^{\text{Asym}}$CDTb, this is because of the position of the linker combining these 2 RBDs is different and because of the formation of the long βBD itself. Thus, the RBD2 in the β-barrel-forming heptamer is located much closer to the protein core as compared to its position in the other heptamer of $^{\text{Asym}}$CDTb or to either heptamer of $^{\text{Sym}}$CDTb. This shift in position is combined with rotation of the entire donut-like structure as the linker is repositioned from a linear to angled orientation (Fig. 4).

Fig. 5. Comparison of the “β-barrel” containing heptamer of $^{\text{Asym}}$CDTb to the analogous heptamer from the protective antigen (PA) of the anthrax toxin. (A). Heptamers from PA of anthrax toxin are superimposed with electron density from the “β-barrel heptamer” observed in the $^{\text{Asym}}$CDTb di-heptamer structure. The RBD in “β-barrel form” of the PA from anthrax toxin were not modeled in the corresponding cryo-EM model and are placed here using alignment with the soluble form of the toxin. (B) Structural comparison of heptameric forms A (Upper) and B (Lower) from $^{\text{Asym}}$CDTb (green) and anthrax toxin (red). Cryo-EM densities are shown for all molecules except anthrax toxin form B for which the 2Fo–Fc map is shown and derived from the corresponding crystal structure.
The Biological Importance of RBD2. Importantly, the second RBD was found to be essential for promoting the di-heptamer assembly in both SymCDTb and AsymCDTb since a C-terminal RBD2 (residues 212 to 751; CDTbΔRBD2), was found to exist as a 7-subunit heptamer and not as a 14-subunit di-heptamer, as determined via SEC-MALS (SI Appendix, Fig. S12). CDTbΔRBD2 also had significantly reduced toxicity in Vero cell killing assays, even at concentrations greater than 10 μM (Fig. 1). Similarly, when the second RBD was isolated (residues 757 to 876), it was found to retain the structure of the intact domain via X-ray crystallography (SI Appendix, Fig. S11) and act as a dominant-negative to block toxicity in Vero cell killing assays (Fig. 1). Specifically, the purified RBD2 protected against a 500-pM dose of the binary toxin in Vero cell killing assays (IC50 = 20 ± 10 nM) (Fig. 1). Additionally, when challenging intact binary toxin with high concentrations of CDTa and CDTbΔRBD2, there was no protection against killing from the intact binary toxin. Taken together, these data show that the unique di-heptamer assembly involving this second RBD has an important role in the binary toxin’s biological activity and represents a domain in active CDTb worthwhile to target via structure-based drug design approaches.

Fig. 6. Detailed structural features of the active CDTb RBDs. (A) Dual calcium binding site located in the N-terminal region of the protein. The coulomb potential maps (i.e., cryo-EM density maps; blue) in both SymCDTb and AsymCDTb resolved 2 Ca2+ ions bound (Ca1, Ca2; green) with Ca1 oxygen ligands from D222/D224/E321/D273/N260(C = O)/E263(C = O), and Ca2+ ligands from D220/D222/D224/E321/D228/I226(C = O). (B) Calcium-binding site located in the β-sandwich domain of RBD1. The RBD1 of CDTb is shown in blue, superimposed with the structure of the β-sandwich from Clostridium thermocellum xylanase Xyn10B used here as an example of Ca2+-binding CBM domain (green). (C) Calcium is required for stability of the isolated RBD1. The 1H,15N-HSQC spectra of RBD1 are illustrated in the absence (blue) and presence (red) of 6 mM CaCl2. A large number of the correlations, in the absence of Ca2+ (blue) were absent due to exchange-broadening or very strong (marked by “x”) consistent with this construct being “unfolded” in the absence of Ca2+. Upon Ca2+ addition, the backbone and sidechain (i.e., for R668ε) correlations appeared and were highly dispersed, consistent with the RBD1 domain folding in a Ca2+-dependent manner. Labeled are resonance assignments for 1H,15N correlations (in red) that are fully correlated with their corresponding 13Cε and 13Cβ chemical shift values, along with 96% of C shifts, and 93% of the side-chain shift values from triple resonance heteronuclear NMR data. Six other correlations were not assigned (marked with an asterisk, *) due to a complete lack of interresidue correlations in the triple-resonance NMR spectra; perhaps some of these unassigned correlations arise from the 6-residue His-Tag used for purifying this domain. Nine other observable correlations (red; labeled with an “x”) were not assigned, even in the presence of Ca2+, and remain disordered based on their narrow line shape and high intensity. Similarly, 19 15N-1H correlations (in blue) for residues of RBD1, in the absence of Ca2+, were not be readily assigned due to their intrinsically disordered state. (D) The predicted secondary structure of RBD1 in the presence of Ca2+ is predominantly β-strands and consistent with that of RBD1 observed in the cryo-EM structures (Fig. 3 and SI Appendix, Fig. S10), and is comprised of 9 β-strands spanning residues: K615-N621; Y625-N626; G645-P659; K667-D677; S683-A690; E693-P700; T705-T714; N720-G727, and Y732-N742.
Possible Biological Role of the Active CDTb Di-heptamer. The core domain structures of each heptamer unit can be predicted based on similarities in sequence to other binary toxins in the 3 heptamerization domains and in the βBD, including 41% sequence identity to the corresponding regions of the anthrax protective antigen. Importantly, however, the first RBD, has no sequence similarity to the corresponding RBD of anthrax toxin, and the anthrax toxin lacks sequence resembling the second RBD of active CDTb altogether. Based on the pre-entry crystal structure of anthrax-protective antigen, it was anticipated that activated CDTb would also be heptameric in structure, particularly since the activity of CDTa:CDTb ratio was optimal at a 1:7 stoichiometry. Thus, the discovery of not 1 but 2 unique di-heptamer structures for activated CDTb was very surprising. Demonstrating the exact nature of the evolutionary advantage conveyed by heptamer dimerization is beyond the scope of this study, but several possibilities can be outlined. The most intriguing of these is that preforming the β-barrel heptamer unit conformation, as found in SymCDTb, may facilitate toxin activity by accelerating its insertion into the membrane and that this process may be facilitated by binding membrane and the host cell receptor. The second heptamer unit in this scenario, the non-β-barrel heptamer, could play the role of a “cap” or “sheath,” protecting and stabilizing the pore-forming heptamer of AsymCDTb, as may be needed to increase its half-life in vivo. However, these and other studies in vivo remain to be completed before such conclusions can be made with certainty.

Another surprising discovery is that a second di-heptamer structure was identified, SymCDTb, and this structure also needs to be considered in mechanistic terms. The essential stabilizing element of this structure is again the central donut-shaped tetradecamer that occurs via intermolecular contacts involving the second RBD. While the biological role of the symmetric structure is not fully understood in vivo, it may facilitate binding to CDTa, membrane, or the host cell receptor. It needs to be pointed out that all of the biophysical data (SEC-MALS, SAXS, AUC) indicated that monomeric CDTb (75 kDa) is still the major species (95 ± 2%) in solutions of activated CDTb, with the 14-subunit oligomer (1.0 MDa) detected at lower levels (≤4 to 6%) (Fig. 2). This result, together with a shifting of this equilibrium toward the 1.1 MDa oligomer via CDTa addition (Fig. 2), is suggestive that monomeric and di-heptameric forms of activated CDTb are in a dynamic equilibrium. With this in mind, interconversion between SymCDTb and AsymCDTb has 2 possible pathways, 1 direct (SymCDTb ↔ AsymCDTb), which was modeled here via normal mode analyses calculations (SI Appendix, Figs. S13–S15), and another via dissociation into the monomeric form (SymCDTb ↔ monomer ↔ AsymCDTb); however, these mechanistic possibilities remain to be established quantitatively as a function of active CDTb concentration. Importantly, though, there was no evidence for a heptameric state of active CDTb, and this region of the toxin could also be a legitimate and unique therapeutic targeting site. Finally, the structures

Fig. 7. Large scale structural features of activated CDTb. (A) Cross heptamer hydrophobic interactions between the tip of the βBD in barrel conformation and HD2 for AsymCDTb. The non-β-barrel (red) and the β-barrel (green) heptameric assemblies of the di-heptamer are distinguished in different colors for clarity. (B and C) Changes in the size of the pore formed by guanine residues (F455) in 2 heptameric forms of the CDTb – AsymCDTb (B) and SymCDTb (C). In B and C, the color scheme is as defined in Fig. 3 with the phenylalnine residues comprising the pore shown in gray.

Summary. No therapeutic is approved by the Food and Drug Administration to target the C. difficile binary toxin in vivo (12). To address this unmet medical need, vulnerabilities in CDT were identified here by solving the structures of SymCDTb and AsymCDTb (Fig. 3). One such targeting strategy is clearly different from methods available to target other heptameric toxins, such as the protective antigen of the anthrax toxin. This involves the unexpected dimerization of 2 heptameric assemblies via the second RBD of the C. difficile binary toxin, which is missing in the anthrax toxin heptamer. This finding makes RBDB2 a particularly promising region of CDT to target via rational drug design methods. Similarly, this study revealed a Ca2+-binding site in the first RBD of active CDTb, and this region of the toxin could also be a legitimate and unique therapeutic targeting site. Finally, the structures
of active CDTb will contribute next to answering important new questions regarding the molecular mechanism of the C. difficile toxin, which can also benefit drug discovery. For example, understanding how the active binary toxin complex assembles and dissociates, how it binds host membrane, how Ca\(^{2+}\) ions affect its function, and how it enters host cells via receptor-mediated processes will certainly build upon these 2 unique and foundational structures reported here for active CDTb.

A second important component of this work is the synergistic approach to structural characterization of active CDTb. Several structural and biophysical methods were employed that provided a multifaceted examination of the problem. Cryo-EM is the nexus of this work as it provided the initial discovery of the active CDTb di-heptamer in 2 different conformations, even for a low percentage of the total protein (4 to 6%). Knowledge of these structural assemblies from cryo-EM then allowed for resolving phasing issues in the crystal structure determination, which provided feedback regarding Ca\(^{2+}\)-binding site in the first RBD of active CDTb. Furthermore, NMR techniques were employed to indicate that Ca\(^{2+}\)-binding to this first RBD is unique to CDT, and it is likely important for its stability. Additionally, the ability to detect multiple conformations for activated CDTb in solution by cryo-EM enhanced the analysis of the SAXS data, which originally provided impetus for considering higher oligomerization states, as the radius of gyration data were inconsistent with a heptamer models based on homology. Even initial models provided an improved fit once conformational heterogeneity was included in the analysis. SAXS data also confirmed that the dual conformation is present in solution and is not an artifact of freezing procedure employed in preparing cryo-EM samples.

Finally, other biophysical techniques (SEC-MALS and AUC) for characterizing size distributions in solution indicated that a significant amount of monomeric protein is present in the solutions used for these structural studies of activated CDTb (>90%). Of note, the structural methods employed here are all insensitive to this for different reasons. Cryo-EM analysis is based on picking particles in the micrographs and thus is dominated by larger clearly discernible megadalton size di-heptamer. In SAXS, larger particles dominate scattering intensity with the detection of smaller monomers being negligible. X-ray crystallography resolves structures that crystallize. In this case, it is remarkable that a conformation that probably represents no more than 2% of the protein particles is the 1 that crystallized, upending the traditional notion that highly concentrating monodisperse protein is a prerequisite of a successful crystal structure determination. Without knowledge of the protein size distribution (>90% monomer in the absence of CDTa) under the conditions of this structure-determination work, a starkly different picture may have arisen for how to describe the transition between the 2 di-heptamer conformations (SI Appendix, Figs. S13–S15) (SymCDTb ↔ AsymCDTb), but these data forced consideration that the conversion between AsymCDTb and SymCDTb could be mediated by an oligomer assembly/disassembly mechanism from the monomeric form (i.e., SymCDTb ↔ monomer ↔ AsymCDTb), which can be affected by other components of the binary toxin complex (i.e., CDTa, membrane, host cell receptor). Furthermore, since monomeric pro-CDTb is not toxic, it opens up yet another therapeutic possibility, as it suggests that fully assembled and active CDTb is in active equilibrium, which may be potentially shifted by small-molecule inhibitors or biologics. Thus, capitalizing on such a multifaceted approach to molecular characterization, it was shown that the resulting picture from the multiple methods is more than the sum of its parts, particularly for large macromolecular assemblies, such as active CDTb (>1 MDa). In summary, the individual structural methods (cryo-EM, X-ray crystallography, NMR) provide phenomenal insights on their own, but they become even more powerful when used together and when combined with other biophysical techniques.

### Materials and Methods

#### Protein Expression and Purification

Active CDTb was expressed and purified as described in ref. 12. Briefly, full length pro-CDTb was expressed in an insect cell–baculovirus system and purified using affinity chromatography. To obtain the active protein, the N-terminal activation domain was proteolytically removed using chymotrypsin and purified by SEC. Full-length CDTa and several truncated constructs of CDTb (RBD1, RBD2, and a construct lacking the second RBD, CDTb\(^{\Delta RBD2}\)) were overexpressed in Escherichia coli and purified to homogeneity by combination of affinity and SEC methods as described in SI Appendix.

#### Vero Cell Activity Assay

Briefly, Vero cells incubated in presence of binary toxin were quantified for F-actin using fluorescently labeled phalloidin to determine toxicity. Further details are provided in SI Appendix.

#### Cryo-EM

Purified and active CDTb was placed on holey gold grids with an additional thin layer of carbon on top, blotted, and flash-frozen in liquid ethane using FEI Vitrobot IV. Grids were inspected and electron micrographs collected on FEI Titan Krios at 300K equipped with Gatan K2 Summit direct electron detector. Multiple iterative rounds of 2D/3D classification resulted in identification of 2 distinct protein conformations for which the density maps were refined with Bayesian particle polishing and CTF refinement with Relion 3.1 (61) and refined with phinex.refine (62). Further details are provided in SI Appendix.

#### X-Ray Crystallography

Crystallization conditions were found for all of the structures via sparse matrix robotic screening. Standard techniques of cryoprotection were used and experimental diffraction data were collected at the Stanford Synchrotron Radiation Lightsource (SSRL). Structures were solved by molecular replacement using Phaser (61) and refined with phenix.refine (62). Further details are provided in SI Appendix.

#### NMR Spectroscopy

A 2D \(^{15}\)N-edited HSQC of 0.5 mM RBD1 in 15 mM Heps (pH 7.0), 150 mM NaCl, 10% D\(_2\)O was collected at 950 MHz, 25 °C. Minimal residues appeared with high noise. 2.3 mM Ca\(^{2+}\) was added and the \(^{15}\)N-edited HSQC was collected under the same conditions. The spectrum was no longer exchange-broadened and lacking dispersion and the number of correlations noticeably increased. The Ca\(^{2+}\) concentration was raised to 6 mM and the spectrum improved further with no additional changes at higher Ca\(^{2+}\) concentrations (>12 mM).

#### Biophysical Techniques

Biophysical characterization of activated CDTb included SAXS, SEC-MALS, and AUC. Experimental details for these techniques are described in SI Appendix.

### ACKNOWLEDGMENTS

Funding from the Center for Biomolecular Therapeutics (D.J.W.), the Maryland Center for Advanced Molecular Analytics (D.J.W. and E.P.), and the University of Maryland School of Medicine are acknowledged. A.D.G acknowledges funding from the NIH (Grant R35GM133598) and the City University of New York. Funding to support the Inductively Coupled Plasma-Mass Spectrometry analyses was from the National Science Foundation (CHE 1708732). Use of the SSR, SLAC National Accelerator Laboratory, is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract DE-AC02-76SF00515. The SSR Structural Molecular Biology Program is supported by the Department of Energy Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of National Institute of General Medical Sciences or NIH.

1. L. L. McDonald et al., Clinical practice guidelines for Clostridium difficile infection in adults and children: 2017 update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). Clin. Infect. Dis. 66, e1–e48 (2018).
2. S. H. Cohen et al.; Society for Healthcare Epidemiology of America; Infectious Diseases Society of America, Clinical practice guidelines for Clostridium difficile infection in adults; 2010 update by the Society for Healthcare Epidemiology of America (SHEA) and the Infectious Diseases Society of America (IDSA). Infect. Control Hosp. Epidemiol. 31, 431–455 (2010).
3. K. A. Brown, N. Kanafer, N. Daneman, D. N. Fisman, Meta-analysis of antibiotics and the risk of community-associated Clostridium difficile infection. Antimicrob. Agents Chemother. 57, 2326–2332 (2013).
4. K. Z. Vandacka, K. K. Trigikidis, E. Boukouvala, M. E. Falagas, Clostridium difficile infection following systemic antibiotic administration in randomised controlled trials: A systematic review and meta-analysis. Int. J. Antimicrob. Agents 48, 1–10 (2016).
5. P. L. Stabler et al., Comparative genome and phenotypic analysis of Clostridium difficile 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol. 10, R102 (2009).
