New tools and insights in physiology and chromosome dynamics of Clostridioides difficile

Oliveira Paiva, A.M.

CHAPTER 3

The C-terminal domain of *Clostridioides difficile* TcdC is exposed on the bacterial cell surface

Ana M. Oliveira Paiva[1,2]

Leen de Jong[1]

Annemieke H. Friggen[1,2]

Wiep Klaas Smits[1,2]

Jeroen Corver[1]

[1] Department of Medical Microbiology, Section Experimental Bacteriology, Leiden University Medical Center, Leiden, The Netherlands

[2] Center for Microbial Cell Biology, Leiden, The Netherlands

Published in Journal of Bacteriology

DOI: 10.1128/JB.00771-19

Abstract

Clostridioides difficile is an anaerobic gram-positive bacterium that can produce the large clostridial toxins, Toxin A and Toxin B, encoded within the pathogenicity locus (PaLoc). The PaLoc also encodes the sigma factor TcdR, that positively regulates toxin gene expression, and TcdC, a putative negative regulator of toxin expression. TcdC is proposed to be an anti-sigma factor; however, several studies failed to show an association between *tcdC* genotype and toxin production. Consequently, TcdC function is not yet fully understood. Previous studies have characterized TcdC as a membrane-associated protein with the ability to bind G-quadruplex structures. The binding to the DNA secondary structures is mediated through the OB-fold domain present at the C-terminus of the protein. This domain was previously also proposed to be responsible for the inhibitory effect on toxin gene expression, implicating a cytoplasmic localization of the OB-fold.

In this study, we aimed to obtain topological information on the C-terminus of TcdC and demonstrate that the C-terminus of TcdC is located extracellularly. In addition, we show that the membrane association of TcdC is dependent on a membrane-proximal cysteine residue and mutating this residue results in release of TcdC from the bacterial cell. The extracellular location of TcdC is not compatible with direct binding of the OB-fold domain to intracellular nucleic acid or protein targets and suggests a mechanism of action that is different from characterized anti-sigma factors.

Importance

Transcription of *C. difficile* toxins TcdA and TcdB is directed by the sigma factor TcdR. TcdC has been proposed to be an anti-sigma factor. The activity of TcdC has been mapped to its C-terminus and the N-terminus serves as a membrane anchor. Acting as an anti-sigma factor requires a cytoplasmic localization of the C-terminus of TcdC.

Using cysteine accessibility analysis and a HiBiT-based system, we show that the TcdC C-terminus is located extracellularly, which is incompatible with its role as an anti-sigma factor. Furthermore, mutating a cysteine residue at position 51 results in release of TcdC from the bacteria. The use of the HiBiT^{opt} system for topology determination of membrane proteins is a valuable tool, increasing the range of available systems to tackle important aspects of the *C. difficile* development.

Introduction

Clostridioides difficile (Clostridium difficile) [1] is an opportunistic pathogen that can cause disease in individuals with dysbiosis of the gut microbiota [2]. *Clostridioides difficile* infection (CDI) incidence has increased worldwide and leads to a broad spectrum of symptoms, from mild diarrhoea to toxic megacolon, and even death [3].

Several factors contribute to the progression and severity of CDI [2,4]. *C. difficile* is a Gram-positive anaerobic bacterium that has the ability to form spores, which allows for dissemination and colonization [2]. The main virulence factors are the large clostridial toxins that induce damage to the epithelial cells and lead to an inflammatory response that underlies the symptoms of CDI [2,3,5].

C. difficile strains have been found to produce up to three toxins: Toxin A (TcdA), Toxin B (TcdB) and binary toxin (CDT) [5,6]. Toxins A and B are encoded by genes *tcdA* and *tcdB*, respectively, located on a 19.6 kb chromosomal region termed pathogenicity locus (PaLoc) [7]. TcdA and TcdB are glucosyltransferases and once translocated to the cytosol of the intestinal epithelial cells, start a cascade of events that can eventually lead to cell death [2,5]. CDT, encoded by the *cdtA* and *cdtB* genes, is an ADP-ribosylating toxin that acts on the actin cytoskeleton [8].

The PaLoc contains at least 3 additional genes that appear to be involved in the regulation of the expression or function of the large clostridial toxins: *tcdE*, *tcdR* and *tcdC* [5,6]. TcdE is a putative holin-like protein, thought to be involved in toxin secretion, however, its exact role is still unclear [9]. TcdR is an RNA polymerase sigma factor that acts as the positive transcriptional regulator of *tcdA*, *tcdB* and *tcdE* and also positively regulates its own expression. A direct interaction between TcdR and RNA polymerase allows the recognition of the target promoters and activates expression [5,10]. Expression of *tcdR,* and consequently *tcdA* and *tcdB,* is influenced by different stimuli, such as temperature, nutrient availability and medium composition [11-14].

Analysis of gene transcription by quantitative PCR has shown that while the expression of *tcdA*, *tcdB*, *tcdE* and *tcdR* is low during exponential phase, these strongly increase upon entering stationary phase [15]. In contrast, *tcdC* was found to be highly expressed during exponential phase but to decrease in stationary phase [15]. Similar profiles were shown at the protein level, where levels of TcdC were higher in the exponential growth phase [16]. Together, these data suggested that TcdC might act as a negative regulator of toxin transcription. However, several other studies did not find a decrease in *tcdC* transcription in stationary

phase but rather showed a constant expression level during the stationary growth phase [13,17,18].

Likewise, the association between toxin expression and *tcdC* gene variants is subject of debate. Increased virulence in epidemic strains was thought to be caused by deletions and frameshift mutations in *tcdC,* leading to a severely truncated non-functional protein and presumably higher toxin titers as a consequence [19,20]. In support of this, it was shown that introduction of a plasmid-based copy of the wild type *tcdC* gene in strain M7404 (PCR ribotype 027, carrying a truncated *tcdC*) resulted in decreased virulence in hamsters [20]. However, mutations in the *tcdC* gene of clinical isolates did not predict the activity of toxins A and B [18,21]. Moreover, several studies failed to observe a relation between toxin gene expression and *tcdC* genotype. Restoration of chromosomal *tcdC* of outbreak strain R20291 (PCR ribotype 027) to wild type did not result in altered toxin expression [22] and toxin expression in *C. difficile* 630Δ*erm* and an isogenic *tcdC* ClosTron mutant showed no significant differences in toxin levels [17].

Previous studies have characterized the domain structure of TcdC [16,23]. TcdC is a 26 kDa dimeric protein that contains an N-terminal transmembrane region (residues 30-50), that allows its anchoring to the cell membrane, a coiled-coil dimerization domain and a C-terminal functional domain [10,23]. Using surface plasmon resonance experiments, purified full-length TcdC was shown to interact with *E.coli* core RNA polymerase and prevented the formation of the active holoenzyme TcdR-RNA polymerase [10]. Overexpression of *C. difficile tcdC* in the heterologous host *Clostridium perfringens* results in repression of TcdR-driven transcription from the *tcdA* promoter, and the C-terminal domain of TcdC was sufficient for this activity [10]. However, it is not clear if TcdR and TcdC are in close proximity inside the bacterial cell.

Due to lack of structural characterization of TcdC homologues, computational analysis was used to build a structural model of the C-terminal domain of TcdC. This modelling suggested the domain adopts a dimeric, ssDNA-binding OB-fold (Oligonucleotide/Oligosaccharide Binding fold) [23]. TcdC is capable of binding to ssDNA G-quadruplexes *in vitro*, but considering the paucity of these structures in the genome sequence of *C. difficile*, G-quadruplexes might mimic an alternative TcdC binding partner [23].

It is clear that further studies are required to understand TcdC binding partners and their function in transcriptional repression. The prevailing model is that TcdC functions as an anti-sigma factor, whose activity depends on cytosolic localization of the C-terminal OB-fold domain. However, no topological information of the C-terminal domain has been demonstrated to date.

In this study, we aimed to determine whether the C-terminal domain of TcdC is cytosolic or surface exposed and evaluated a codon-optimized version of the HiBiT Extracellular Detection System (HiBiT^{opt}) as a valuable addition to the molecular tools to study *C. difficile*. We find that the C-terminal domain of TcdC is located extracellularly and show the value of the HiBiT^{opt} system for topology studies of *C. difficile* proteins. In addition, we show that membrane association of TcdC is dependent on a membrane-proximal cysteine residue and that mutating this residue results in release of TcdC from the bacterial cell.

Results

In-silico prediction of TcdC topology suggests an extracellular location of the C-terminal domain.

To analyze the topology of *C. difficile* TcdC (CD0664 from *C. difficile* 630) we first analyzed the protein sequence (Uniprot ID: Q189K7) using three different prediction algorithms: TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0) [24], TOPCONS 2.0 (http://topcons.cbr.su.se/) [25] and SignalP 5.0 (http://www.cbs.dtu.dk/services/SignalP/) [26].

TMHMM 2.0 [24] predicts a transmembrane helix of around 16 residues (residues 31-46) with moderate probability. Residues 1-13 are predicted to be inside of the cell (0.63 probability) whereas the C-terminal region (coiled-coil and OB-fold domains) is predicted to be outside of the cell (probability >0.8) (Fig 1A). The consensus of TMHMM 2.0 1-best algorithm predicts TcdC to be extracellular (Fig. 1A, pink bar). TOPCONS 2.0 [25], which identifies regions with a low free energy difference (ΔG), similarly suggests the presence of a transmembrane helix. The TOPCONS 2.0 consensus prediction is an intracellular N-terminal domain (residues 1 – 26, Fig. 1B, red bar), a transmembrane helix (residues 27 – 46, Fig. 1B, grey bar) and an extracellular C-terminal region that encompasses the dimerization and OB-fold domains (residues 47 – 232, Fig. 1B, blue bar).

We also investigated the presence of a potential signal peptide in the TcdC amino acid sequence through SignalP 5.0 [26]. However, no known signal peptide was identified, suggesting TcdC remains tethered to the membrane (Fig. 1C).

Though the reliability of the predictions is relatively low, both TMHMM and TOPCONS support the presence of the transmembrane helix (Fig. 1), consistent with previous observations [10,16]. Strikingly, both methods suggest that the TcdC C-terminus is located outside of the cell. As this would be incompatible with a role for the OB-fold domain in sequestering TcdR or

repression of TcdR-mediated transcriptional activation, we set out to obtain topological information on the C-terminal domain in *C. difficile*.

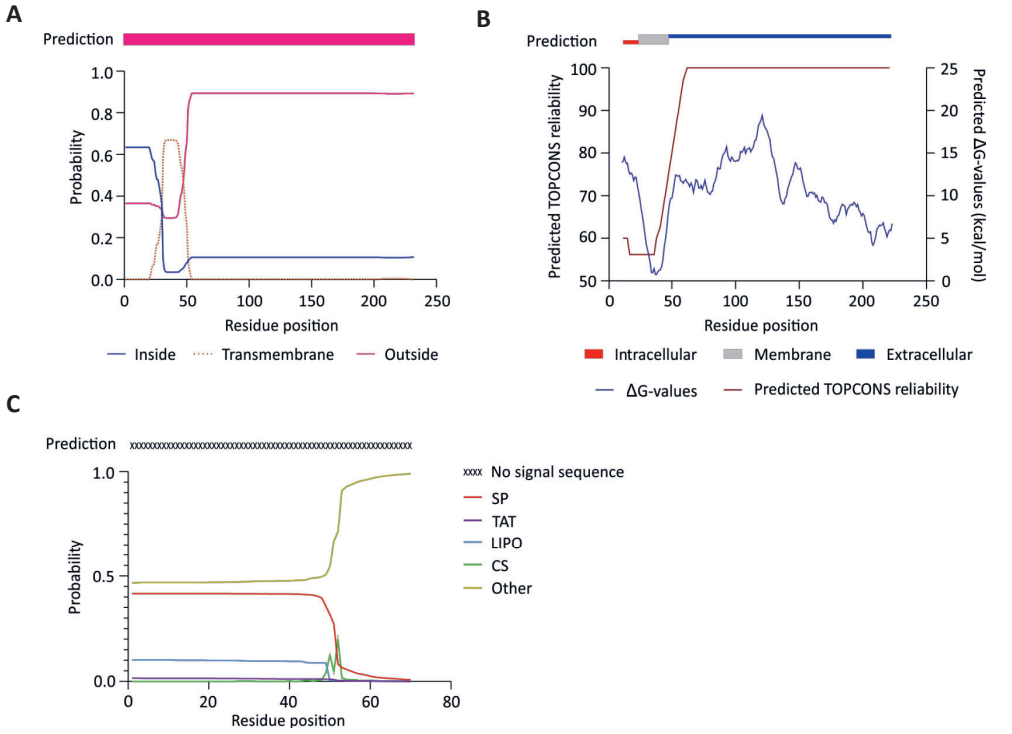


Fig. 1 - Prediction of a transmembrane helix in TcdC. A) Output from the prediction by TMHMM 2.0 software [24] through 1-best algorithm (pink bar) and probability plot: inside the cell (blue line), transmembrane region (orange dotted line) and outside the cell (pink line). **B)** Prediction by TOPCONS software [25], with consensus in residues 1-26 inside the cell (red box), a transmembrane helix (residues 26-46, grey box) and residues 47-232 on the outside of the cell (blue box). TOPCONS reliability score (brown line) and predicted ΔG-values for each residue (blue line) are shown. **C)** Output from the SignalP 5.0 [26] web-server for the TcdC amino acid sequence. No signal sequence was detected (X). Probabilities of signal peptides presence from the systems Sec (SP, red line), Tat (purple line), and lipoprotein (LIPO, blue line) are shown. Predicted cleavage site score (CS, green line) and no signal sequence probability is depicted (OTHER, light green line).

TcdC is accessible for extracellular cysteine labelling

We analyzed whether cysteines natively present in TcdC are exposed to the extracellular environment in a manner similar to Substituted Cysteine Accessibility Method (SCAM). SCAM subjects cysteine residues present in the protein of interest to chemical modification with the thiol-specific probe N-(*3*-maleimidylpropionyl) biocytin (MPB), that has a low membrane permeability. The probe forms a stable, non-hydrolyzable bond with the thiol-group of a cysteine residue, resulting in the biotinylation of the protein. At low concentrations of MPB,

exclusively extracytoplasmic (surface exposed) cysteines are labelled, providing topological information about the labelled protein [27]. A typical SCAM experiment relies on immunoprecipitation of protein (using an antibody specific for the protein of interest), detection of immunoprecipitated protein (using a second antibody, directed at a tag on the protein of interest), and verification of labelling with the MPB (using anti-biotin antibodies).

We introduced a C-terminally 3xmyc-tagged TcdC expression construct (TcdC-3xmyc, Fig. 2A), with an otherwise native protein sequence, under the control of the inducible promoter P_{tet} [28], that can be precipitated with anti-TcdC (α-TcdC) antibody and detected using anti-myc antibodies (α-myc). We affinity purified a previously generated TcdC antibody [17] and verified its specificity on *C. difficile* lysates by immunoblotting. The TcdC-3xmyc construct was induced and samples before and after induction were analyzed. Only in the induced samples a band migrating at the approximate molecular weight of TcdC was observed (38 kDa, Fig. 2B, red arrow), suggesting that native levels of TcdC under non-inducing conditions are below our limit of detection in this assay. Though the predicted molecular weight of the TcdC-3xmyc protein is 26 kDa, the observed molecular weight has been reported as 37 kDa [10,23], due to unknown reasons. Several bands were detected with a lower molecular weight than expected (Fig. 2B). The fact that these are only present when TcdC-3xmyc is induced (Fig. 2B) suggests possible alternative forms of the protein. Nevertheless, the apparent specificity of the anti-TcdC antibody allowed further analysis.

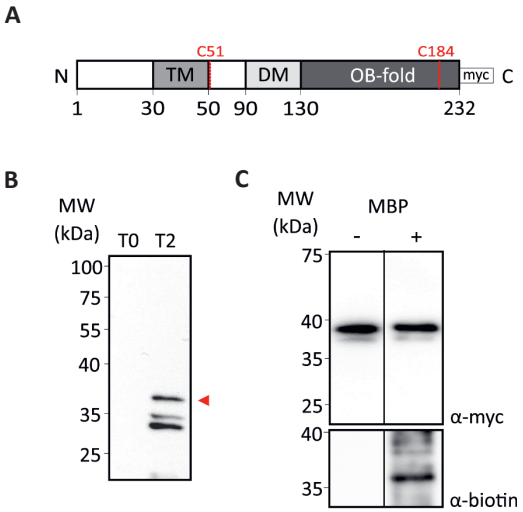


Fig. 2 - Mapping the location of the TcdC C-terminus with cysteine accessibility analysis. A) Schematic representation of the C-terminally 3xmyc-tagged TcdC construct used for the cysteine accessibility analysis. The different domains of TcdC are represented: transmembrane domain (TM, grey box), the

dimerization domain (DM, light grey box) and the OB-fold (dark grey box). The 3xmyc-tag is represented as a white box, the cysteines residues present on TcdC are represented (zigzag red line). **B)** Western-blot analysis of α-TcdC antibody specificity in *C. difficile* 630Δ*erm* lysates harbouring pLDJ1 (P_{tet}-*tcdC-3xmyc*), before (T0) and after induction with 200ng/ml anhydrotetracycline for 2 hours (T2). Full-length TcdC is indicated with a red arrow. **C)** Cysteine labelling analysis of the different TcdC-3xmyc constructs. The strain harbouring the C-terminally 3xmyc-tagged TcdC construct (38 kDa) was induced for 2 hours. Samples were collected and either not treated with MPB (-) or treated with 1 mM MPB (+). Samples were immunoprecipitated and immunoblotted with α-myc for TcdC-3xmyc protein detection (upper panel), and α-biotin for detecting biotinylated proteins (bottom panel). Cysteine biotinylation of TcdC-3xmyc was observed.

TcdC contains 2 endogenous cysteines; one at position 51, right after the predicted transmembrane domain (residues 30 to 50) and another one at position 184, located in the predicted OB-fold (Fig. 2A). To evaluate the cysteine labelling on the native protein, we assessed the biotinylation of TcdC-3xmyc (Fig. 2C). The signal in the anti-biotin (α-biotin) Western blot suggests that one, or both, of the native cysteine residues in TcdC-3xmyc, is accessible for labelling by MPB (Fig. 2C). We could not perform a full SCAM analysis [27], as mutation of the native cysteines that is necessary for such an analysis (see further below), resulted in inconsistent and low-level expression of cell-associated TcdC.

Nevertheless, as only a single membrane spanning region is predicted this result suggests that the TcdC C-terminus, where both cysteines are present, is located extracellularly.

HiBiT^{opt} assay for *C. difficile* confirms the extracellular location of the TcdC C-terminus

We sought to confirm the results of the cysteine accessibility assay in an independent experiment. Previously, we have successfully used luciferase reporter assays to assess promoter activity and *in vivo* protein-protein interactions in *C. difficile* [29,30]. Here, we extend the luciferase toolbox for *C. difficile* by validating an adaptation of the Nano-Glo® HiBiT Extracellular Detection System (Promega) [31].

Similar to the SmBit, the HiBiT tag is a small 11 amino acid peptide that binds to a larger subunit called LgBiT to reconstitute a functional luciferase [30-32]. However, in contrast to the SmBit, HiBiT has been engineered for high affinity for the LgBiT subunit [32]. Due to its molecular weight (19 kDa), extracellularly added LgBiT cannot enter the cell. Thus, a luminescent signal in the presence of the substrate furimazine is only observed if the HiBiT subunit is accessible from the extracellular environment [31].

To apply this system for detection of *C. difficile* protein topology we constructed several controls carrying codon-optimized C-terminal HiBiT (HiBiT^{opt}) tags, as schematically represented in Fig. 3A. As a positive control for the detection of extracellular proteins, the

Sortase B (SrtB) protein was selected (SrtB-HiBiT^{opt}). Sortases are membrane-anchored enzymes which catalyze the cleavage and transpeptidation of specific substrates and thereby facilitate their attachment to the cell wall [33]. The genome of *C. difficile* strain 630 (and also its derivative 630Δ*erm*) has a single sortase, SrtB, present at the *C. difficile* cell wall [34,35]. The localization of SrtB and its substrates at the *C. difficile* cell surface makes SrtB a suitable candidate for the extracellular detection of the reconstituted luminescent signal. As negative control, the HupA protein was used (HupA-HiBiT^{opt}). This protein is a cytosolic DNA binding protein that is not secreted to the extracellular environment and thus should not be accessible to the LgBiT subunit [30]. All the constructs were placed in a modular vector under the control of the anhydrotetracycline (ATc) -inducible promoter P_{tet} [28]. As observed in other bioluminescence assays [29,30] a background signal is detected from non-induced cells (Fig. 3B, T0), which is comparable to that of a medium only control (17872.8 ± 4397.7 RLU/OD; data not shown). As expected, expression of the positive control SrtB-HiBiT^{opt} leads to a 2-log increase of the luminescence signal after 45 minutes of induction ($3.2x10^6$ ± $2.5x10^5$ RLU/OD, T1, Fig. 3B). No significant increase of the luminescent signal was detected in the cells expressing the negative control HupA-HiBiT^{opt}, confirming that LgBiT does not enter *C. difficile* cells. The lack of signal is not due poor induction, as a clear signal is visible in the lysates of the induced samples for both the SrtB-HiBiT^{opt} and HupA-HiBiT^{opt}, at the expected molecular weights of 26 kDa and 12 kDa, respectively (Fig. 3C). To confirm that the obtained luciferase signals were derived from cells with an intact cell envelope, we performed lysis on the same samples. Indeed, after lysis of the cells, a clear increase of the luciferase signals was observed for both the SrtB-HiBiT^{opt} and HupA-HiBiT^{opt} (Fig. 3B, lysed). A modest, but significant, 1-log increase was also observed for SrtB-HiBiT^{opt} ($8.4x10^7$ ± $9.5x10^6$ RLU/OD, lysed, Fig. 3B), which we attribute to enhanced accessibility in SrtB-HiBiT^{opt} for the substrate due to the lysis. However, for cells expressing the HupA-HiBiT^{opt} a significant 4-log increase in the luciferase signal was observed ($1.1x10^8$ ± 2.4 $x10^7$ RLU/OD, lysed, Fig. 3B). These results confirm that LgBiT does not enter *C. difficile* cells and that the HiBiT^{opt} system as employed is suitable for determining the subcellular localization of the C-terminal domain of *C. difficile* proteins.

Next, HiBiT^{opt} was fused to the C-terminus of TcdC (TcdC-HiBiT^{opt}). We observed a 3-log increase in the luminescence signal after 45 min of induction ($3.2x10^7$ ± $3.0x10^6$ RLU/OD at T1, Fig. 3B). No further increase was detected when cells were lysed ($1.4x10^8$ ± $3.7x10^7$ RLU/OD, lysed, Fig. 3B), supporting an extracellular location of the TcdC C-terminus. Through blotting of bacterial lysates and subsequent measuring luminescence on the blot (see materials and methods), expression of the TcdC-HiBiT^{opt} was confirmed by detection of a clear signal at the expected MW of approximately 39 kDa (Fig. 3C). We observed a low-level signal for the non-

induced TcdC-HiBiT^{opt} expression construct, both in the luciferase assay (T0, Fig. 3B) and in the detection of tagged protein (T0, Fig. 3C), which was not observed for SrtB or HupA. As all proteins are expressed from the same promoter, this possibly indicates more efficient translation, and thus higher expression, of TcdC-HiBiT^{opt} under non-inducing conditions. Alternatively, differences in luciferase detection levels might be explained by protein stability or accessibility of the HiBiT^{opt} fusion proteins for the LgBiT subunit, which in turn is affected by the structure and the exact localization of the proteins.

The HiBiT^{opt} experiments indicate that the C-terminus of TcdC is located in the extracellular environment (like SrtB) and not in the intracellular environment (like HupA).

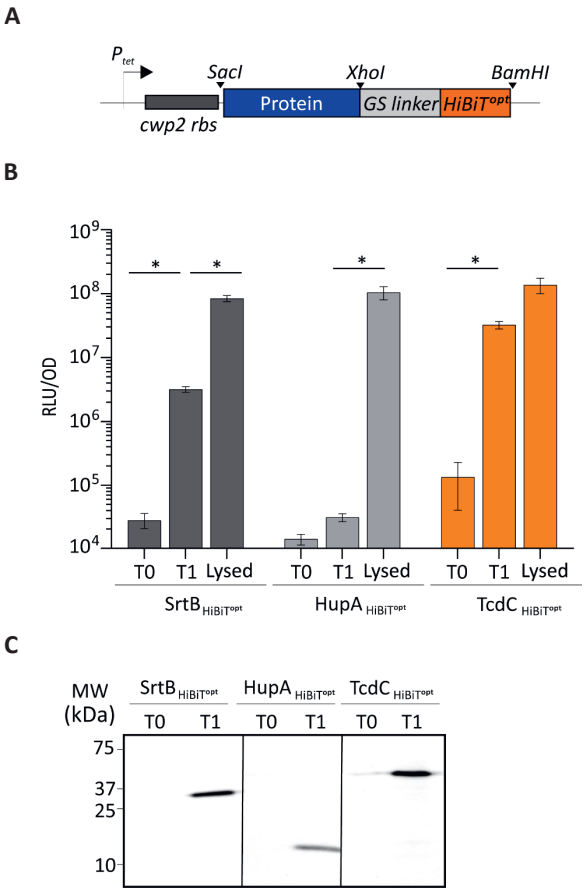


Fig. 3 – Detection of C-terminal HiBiT^{opt} tags. A) Representation of the HiBiT^{opt} modular cassette. The protein of interest (blue box) fused at the C-terminus to the HiBiT^{opt} (orange box) through the GS linker (grey box) are indicated. The positions of used restriction sites are marked (SacI, XhoI and BamHI) and the *cwp2* ribosomal binding site (dark grey box) are represented. **B)** Proteins of interest were C-

terminally fused to a HiBiT protein tag and their expression was induced with 50 ng/mL ATc for 45 minutes. Optical density-normalized luciferase activity (RLU/OD) is shown right before induction (T0), after 45 min of induction (T1) and subsequent lyses of T1 samples (lysed). HiBiT^{opt}-tagged sortase (dark grey bars) and HupA (light grey bars) proteins were used as extracellular and intracellular controls, respectively. TcdC-HiBiT^{opt} associated luciferase activity is displayed in the orange bars. The averages of biological quadruplicate measurements are shown, with error bars indicating the standard deviation from the mean. Significance was defined as higher than *p<0.001 by two-way ANOVA. **C)** Blot detection of HiBiT^{opt}- tagged proteins resolved on a 12% SDS-PAGE. Sample volumes were normalized for optical density of the cultures from which they were derived. Expression of HiBiT^{opt} fused proteins was observed at 0 (T0) and 45 minutes after induction (T1).

The cysteine residue at position 51 is important for membrane association of TcdC

While assessing the extracellular accessibility of the cysteines in TcdC, we planned to perform a classic SCAM analysis that would allow us to determine which one (or both) of the extracellular cysteine residues were labelled and could potentially also be used to confirm the subcellular localization of the N-terminus of the TcdC protein. To this end, we constructed several mutants of TcdC-3xmyc. During the experiments, we observed inconsistent and low-level expression of cell-associated TcdC, when the cysteine residue at position 51 was mutated into a serine residue (Fig. 4A). Likewise, mutation to alanine yielded very low levels of TcdC in *C. difficile* cell lysates, indicating that this effect is not specific to the serine substitution (Fig. 4A). We reasoned that the expression levels of these TcdC mutants were unlikely to differ as a result of a single amino acid change and therefore we analyzed the culture supernatant by immunoblotting using anti-TcdC antibodies to assess whether TcdC was released from the cells. Supernatants of cells, pelleted two hours post-induction, contained significant amounts of TcdC when cysteine 51 was mutated (Fig. 4B). In contrast, TcdC was cell-associated in constructs expressing wild-type TcdC that contained the cysteine 51 residue. Remarkably, TcdC present in the supernatant was of a significantly smaller size than cell-associated TcdC (Fig. 4E), suggesting TcdC may be the subject of a proteolytic event.

To confirm these results, we also mutated the cysteine 51 residue in the TcdC-HiBiT^{opt} into a serine residue, monitored the luciferase activity in the culture supernatant of cells and compared it to the wild-type TcdC-HiBiT^{opt}. Luciferase activity in supernatants of cells expressing TcdC(C51S)-HiBiT^{opt} was approximately 4 fold higher than of cells expressing TcdC-HiBiT^{opt} (Fig. 4D), indicating that TcdC(C51S) was released from the bacterial cells to a greater extent than wild type TcdC. On the other hand, total signal (cells and medium together) was equal for TcdC and TcdC(C51S), showing that the difference in the supernatant was not due to increased expression of TcdC(C51S) (Fig. 4C).

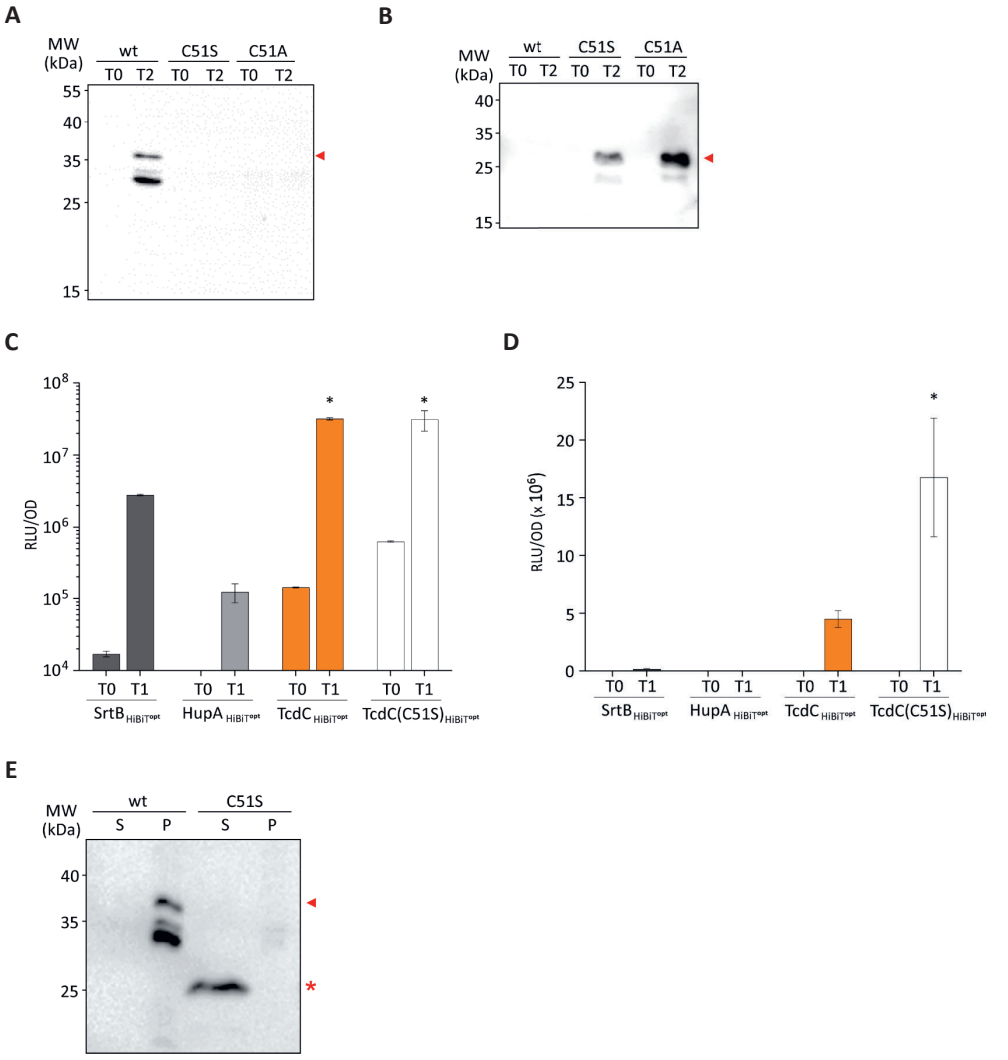


Fig. 4 - TcdC C51S affects membrane localization. A) Western-blot analysis, with α-TcdC, of *C. difficile* 630Δ*erm* cell lysates harbouring pLDJ1 (P_{tet}-*tcdC-3xmyc*), pLDJ2 (P_{tet}-*tcdC(C51S)-3xmyc*) and pJC129 (P_{tet}-*tcdC(C51A)-3xmyc*), before (T0) and after induction with 200ng/ml ATc for 2 hours (T2). TcdC is indicated with a red arrow. **B)** Western-blot analysis, with α-TcdC, of *C. difficile* 630Δ*erm* culture supernatants harbouring pLDJ1 (P_{tet}-*tcdC-3xmyc*), pLDJ2 (P_{tet}-*tcdC(C51S)-3xmyc*) and pJC129 (P_{tet}-*tcdC(C51A)-3xmyc*), before (T0) and after induction with 200ng/ml ATc for 2 hours (T2). Secreted/released TcdC is indicated with a red arrow. **C)** Proteins of interest were C-terminally fused to a HiBiT protein tag and their expression was induced with 50 ng/mL ATc for 45 minutes. Optical density-normalized luciferase activity (RLU/OD) of the culture (cells plus medium) is shown right before induction (T0), after 45 min of induction (T1). HiBiT^{opt}-tagged sortase (dark grey bars) and HupA (light grey bars) proteins were used as surface-exposed and intracellular controls, respectively. Luciferase activity with TcdC-HiBiT^{opt} and TcdC(C51S)-HiBiT^{opt} is displayed in orange or white bars, respectively. The averages of biological triplicate measurements are shown, with error bars indicating the standard deviation from the mean.

Significance was defined as *p<0.001 by two-way ANOVA. **D)** Observed luciferase activity (RLU) in supernatants only, from the cells in C). **E)** Comparison of cell-associated and cell-released TcdC. The same samples as in A) and B) were run next to each other for a fair comparison of the size of the proteins. Cell-associated TcdC is indicated with a red arrow, cell-released TcdC is indicated with a red star.

Discussion

The importance of TcdC for regulation of toxin expression is highly controversial. Though the prevailing model suggests that TcdC is an anti-sigma factor with a role as a negative regulator of toxin transcription [19,20], several other studies have found no clear relation between TcdC expression and the toxin levels [17,18,21,22]. Previous biochemical analyses of TcdC revealed that it is membrane-associated and the TcdC C-terminus comprises a dimerization domain and a domain with a predicted OB-fold important for transcriptional repression [10]. However, the localization of the C-terminus of TcdC has not been studied to date and this was addressed in the present study.

In silico analyses of the TcdC amino acid sequence using TMHMM 2.0 [24], TOPCONS [25], and SignalP 5.0 [26] suggest the presence of a transmembrane helix and predicts no high-probability cleavage for a secretion signal for any of the canonical secretion pathways (Fig. 1). The prediction of a transmembrane domain is consistent with the previously described biochemical assays demonstrating association of TcdC with the *C. difficile* membrane [16,23]. Analysis did not reach consensus on the localization of the N-terminus, due to low reliability scores and differences obtained with the prediction methods (Fig. 1). Nevertheless, both TMHMM 2.0 and TOPCONS suggest that the C- terminus of TcdC (residues 50 to 232, Fig. 1) is located on the outside of the cell.

For a preliminary analysis on the C-terminus accessibility, we took advantage of the 2 cysteine residues that are natively present after the transmembrane helix of TcdC (Fig.2A). Biotin labelling of TcdC-3xmyc was detected, suggesting one or both of the native cysteine residues are accessible for labelling by MPB [27] and therefore located in the extracellular environment (Fig. 5).

Our result was confirmed in independent experiments using the HiBiT^{opt} system, which to our knowledge is applied here for the first time in *C. difficile*. These experiments confirm that the C-terminus of TcdC is exposed on the cell surface. Unfortunately, it was not possible to apply the HiBiT^{opt} system to determine the localization of the N-terminus of TcdC. Fusions of HiBiT^{opt} at the N-terminus of proteins with an established intracellular localization resulted in a high extracellular luciferase signal (data not shown). We attribute this to possible N-terminal processing, which could lead to the release and secretion of the small tag. Further

optimization of the HiBiT^{opt} system is essential before the system can be used to assess localization of both protein termini.

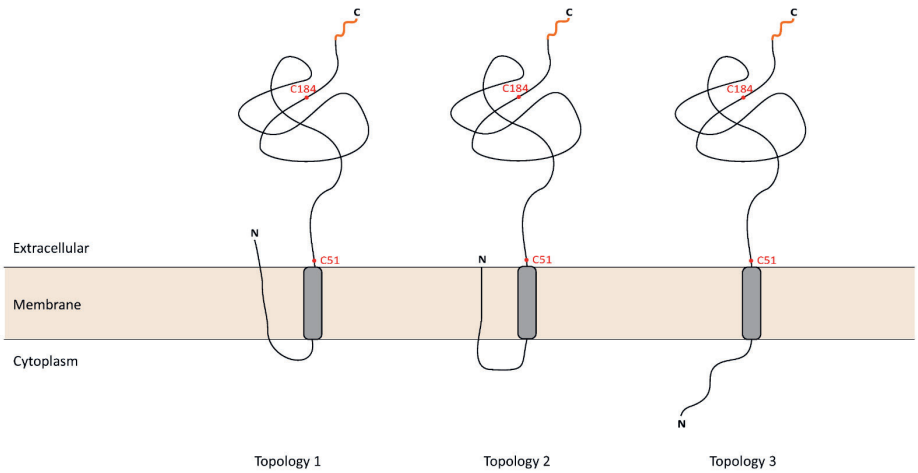


Fig. 5 - TcdC topology models. TcdC is located in the cell membrane with an extracellular C-terminal region. The localization of the 50-aminoacid N-terminal domain of TcdC is unknown. In topology model 1 the N-terminus can cross the cell membrane, exposing the N-terminal domain to the extracellular environment. Another possibility, topology model 2, has the N-terminus present in the cell membrane where it is not accessible from the extracellular or intracellular milieu. Finally, in topology model 3, the N-terminus is present in the intracellular environment of the cell. Cysteines residues used for the cysteine accessibility analysis (red dots) and the C-terminal location of the HiBiT^{opt} tag (orange line) are indicated.

We find that TcdC is released from the cell when a membrane-proximal cysteine residue is mutated in the protein (Fig. 4). As mentioned earlier, TcdC in culture supernatant appeared smaller in molecular weight than cell-associated TcdC, suggesting a possible cleavage event. *In silico* analysis using SignalP suggests that both TcdC(C51S) and TcdC(C51A) are good substrates for signal peptidase (likelihood 0.6671 and 0.6901 with a probability of the cleavage site between S52 and E53 of 0.5323 and 0.4482, respectively, see also Supplemental Fig. S1), whereas TcdC is not (likelihood of 0.4169, no score given for probability). As signal peptidase is located in the extracellular milieu, this provides an additional indication of the extracellular localization of the TcdC C-terminal OB-fold. Alternatively, the decreased size of TcdC(C51S) could be the result of the lack of cysteine-specific post-translational modification. Such a modification could also be responsible for the higher-than-expected molecular weight of wild type TcdC in immunoblotting experiments. Glycosylation is one of the most common post-translational modifications found in several bacteria, particularly at the cell surface [36]

and the higher molecular weight TcdC bands could represent glycosylated forms of the protein. Targeted mass spectrometry-based proteomics might shed light on the nature of these modifications if this is the case. Although our data cannot discriminate between these two hypotheses at this point, we favour the hypothesis that TcdC is cleaved when the C51 is not present as this is supported by the SignalP predictions.

When analyzing the possible release of TcdC mutants, we were unable to detect TcdC by Western-blot in the supernatant of cells expressing TcdC. In contrast, when using the HiBiT^{opt}-tagged TcdC, we were able to detect some luciferase activity in the supernatant, indicating release of TcdC (see Fig. 4D). We can only speculate about this discrepancy. First of all, we cannot compare the sensitivity of the two assays. The fact that we cannot detect TcdC in the Western-blot may be due to the limited sensitivity of the assay. Second, overexpressing of TcdC may lead to some release that can be measured with the highly sensitive HiBiT^{opt} based assay. However, since no significant activity was measured in the supernatants of cells overexpressing Sortase-HiBiT^{opt}, spontaneous release of protein due to overexpression per se does not seem a reasonable explanation for the luciferase activity in the supernatant of cells overexpressing TcdC.

When predicting cleavage of TcdC through SignalP, a low score was given to another possible cleavage system (other than Sec/SPI, TAT/SPI or Sec/SPII). It is possible that part of TcdC is indeed cleaved by another protease, but we cannot speculate about the nature of this protease. Mutating the membrane-proximal cysteine for serine or alanine converts TcdC in a likely substrate for the efficient Sec/SPI system, possibly explaining the increased signal on Western-blot and in the HiBiT^{opt} based assay.

Our results show an extracellular localization of the C-terminus of TcdC, but we could not fully explore the topology of TcdC. Prediction of the TcdC N-terminus was not unambiguous (Fig. 1) and the experimental approaches were unable to conclusively demonstrate the localization of the N-terminus of TcdC due to technical limitations. The protein may therefore adopt different orientations, as represented in Fig. 5. The N-terminus may be located extracellularly (topology 1), embedded in the cell membrane (topology 2) or be exposed in the intracellular environment (topology 3). Thus, a careful further characterization of the TcdC topology is required by alternative means. Recently the Fluorescence Activating and absorption Shifting Tag (FAST) has been used to label proteins and assess protein topology in bacteria [37]. The use of non-permeant substrates allows the localization of exposed proteins on the surface in gram-negative and gram-positive organisms [37]. FAST can be used in anaerobic environments

and was previously used in *C. difficile* [38], which makes it a promising candidate to explore the N-terminal location of TcdC.

Our finding that the TcdC C-terminus is extracellular challenges the prevailing model of TcdC as an anti-sigma factor. Anti-sigma factors generally sequester their cognate sigma factors away from RNAP using substantial cytoplasmic domains [39]. The small N-terminal sequence, that may or may not be intracellular (Fig. 5), is not likely to fulfil this function. Our experiments place the C-terminal domain, that previously was postulated to be responsible for transcriptional repression [10], in a different environment than TcdR and RNAP. One has to wonder whether the OB-fold would ever be in contact with these proteins, as was demonstrated in experiments using purified proteins (outside the context of a *C. difficile* cell) and in heterologous expression systems [10]. Experiments that show repression of TcdR-mediated transcription by co-expression of TcdC were carried out in a heterologous background [10]. It is unclear in this paper what directs the *tcdR* transcription and the observed inhibition of TcdR-mediated transcription by TcdC may be explained by an indirectly lowered level of *tcdR* transcription. In addition, the inhibitory effect of TcdC was also seen when TcdC was expressed without its N-terminal hydrophobic part, which led to the conclusion that the C-terminal domain is responsible for the inhibitory effect. However, removal of a membrane-binding domain will likely result in mislocalization of the protein compared to the wild type situation and does not represent a physiological setup. The fact that this truncated TcdC inhibits TcdR-mediated transcription argues for an aspecific effect of TcdC expression. Lastly, biochemical evidence based on surface plasmon resonance may have been influenced by the fact that full-length TcdC, including the hydrophobic stretch held responsible for the TcdC membrane association, was used. Regardless, it should be noted that an extracellular location of TcdC does not exclude a function as a negative regulator of toxin production, but if so, suggests that it does so through an indirect mechanism.

Our data rather implies binding of the TcdC OB-fold to an extracellular ligand. Bacterial OB-fold proteins have been identified in bacterial genomes and can bind a wide variety of molecules [39,40]. Thus, TcdC might bind extracellular oligonucleotides and/or oligosaccharides. It has previously been shown that the TcdC OB-fold is able to bind G-quadruplex structures, but the physiological relevance of this binding has yet to be determined and it is conceivable that G-quadruplex binding mimics binding of its natural substrate as proposed earlier [23]. In the extracellular environment, TcdC might bind oligosaccharides or extracellular DNA (eDNA) [41-43]. In *Staphylococcus aureus* the sAg-like protein 10 (SSL10) binds to human IgG1 Fc primarily by its N-terminal OB-fold domain and can play a role during *S. aureus* infections [44]. In *Salmonella typhimurium,* the small periplasmic protein YdeI contains an OB-fold domain and

contributes to the resistance to antimicrobial peptides by interaction with the OmpF porin [45]. The VisP protein, a protein from the bacterial oligonucleotide/oligosaccharide-binding fold family also present in S. *typhimurium*, binds to the peptidoglycan sugars and also to the inner membrane protein LpxO, mediating resistance and pathogenesis in *S. typhimurium* [46]. To identify the TcdC ligand(s), a cross-linking and subsequent mass-spectrometry based method with tagged TcdC could be used. In addition, structural studies of TcdC and its ligands could show how the OB fold of TcdC has evolved and to what extend TcdC contributes to downregulation of the large *C. difficile* toxins.

In summary, we developed and applied for the first time a modular vector that allows the C-terminal tagging of proteins with a HiBiT^{opt}, which extends our existing luciferase toolkit [29,30]. This system offers a useful method for the determination of the topology of C-terminal domains of membrane proteins in cells grown under anaerobic conditions without complex processing of samples. Our study indicates an extracellular localization of the C-terminus of TcdC, which is incompatible with its proposed function as an anti-sigma factor. Further studies are required to elucidate the role of TcdC in *C. difficile* development and toxin regulation.

Materials and Methods

Topology prediction

To determine the topology of *C. difficile* TcdC protein (UniProt ID: Q189K7) the amino acid sequence was analyzed by two computer programs for transmembrane and topology assessment: TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0) [24], with extensive output format, with graphics; and TOPCON 2.0 (http://topcons.cbr.su.se/) [25]. The TcdC sequence was analyzed with SignalP 5.0 program, for signal peptide prediction, with long output for gram-positive organisms, (http://www.cbs.dtu.dk/services/SignalP/) [26]. All the results were visualized with GraphPad Prism 8 software (version 8.1.2)

Strains and growth conditions

E. coli strains were grown aerobically at 37°C in Luria Bertani broth (LB, Affymetrix) supplemented with chloramphenicol at 15 µg/mL or 50 µg/mL kanamycin when required. Plasmids (Table 1) were maintained in *E. coli* strain DH5α and transformed using standard procedures [47]. *E. coli* CA434 [48], was used for plasmid conjugation with *C. difficile* strain 630Δ*erm* [34,49].

C. difficile strains grown anaerobically in Brain Heart Infusion broth (BHI, Oxoid), with 0,5 % w/v yeast extract (Sigma-Aldrich), supplemented with *Clostridioides difficile* Selective Supplement (CDSS; Oxoid) and 15 µg/mL thiamphenicol when necessary, at 37°C in a Don Whitley VA-1000 workstation or a Baker Ruskin Concept 1000 workstation with an atmosphere of 10% H_2, 10% CO_2 and 80% N_2. All strains are described in Table 2.

The growth was monitored by optical density reading at 600 nm (OD_{600}).

Table 1 - Plasmids used in this study.

Name	Relevant features *	Source/Reference
pCR2.1-TOPO	TA vector; pMB1 *oriR*; *km amp*	ThermoFisher
pCR2.1TcdC	pCR2.1-TOPO with *tcdC; km amp*	This study
pRPF185	*tetR* P_{tet}-*gusA*; *catP*	[28]
pLDJ1	*tetR* P_{tet}-*tcdC-3xmyc*; *catP*	This study
pLDJ2	*tetR* P_{tet}-*tcdC(C51S)-3xmyc*; *catP*	This study
pAF302	*tetR* P_{tet}-*hupA-hiBiT^{opt}*; *catP*	This study (Addgene ID 137752)
pAP233	*tetR* P_{tet}-*srtB-hiBiT^{opt}*; *catP*	This study (Addgene ID 137753)
pJC111	*tetR* P_{tet}-*tcdC-hiBiT^{opt}*; *catP*	This study
pJC127	*tetR* P_{tet}-*tcdC(C51S)-hiBiT^{opt}*; *catP*	This study
pJC129	*tetR* P_{tet}-*tcdC(C51A)-3xmyc*; *catP*	This study

** amp* – ampicillin resistance cassette, *km* – kanamycin resistance cassette, *catP* – chloramphenicol resistance cassette

Table 2 - *C. difficile* strains used in this study.

Name	Relevant Genotype/Phenotype*	Origin /Reference
630Δ*erm*	*C. difficile* 630Δ*erm*; Erm^S	[34,49]
LDJ1	630Δ*erm* pLDJ1; $Thia^R$	This study
LDJ2	630Δ*erm* pLDJ2; $Thia^R$	This study
AP239	630Δ*erm* pAP233; $Thia^R$	This study
JC267	630Δ*erm* pJC111; $Thia^R$	This study
JC271	630Δ*erm* pAF302; $Thia^R$	This study
JC324	630Δ*erm* pJC127; $Thia^R$	This study
JC326	630Δ*erm* pJC129; $Thia^R$	This study

* Erm^S – Erythromycin sensitive, $Thia^R$ – Thiamphenicol resistant

Strain construction

All oligonucleotides from this study are listed in Table 3. All PCRs were performed on *C. difficile* 630Δ*erm* genomic DNA [34], unless indicated otherwise. Expression vectors are all based on pRPF185 [28]. All DNA sequences in the recombinant plasmids were verified by Sanger sequencing of the region of the plasmid encompassing the inserted fragment and the full anhydrotetracycline-inducible promoter.

Table 3. Oligonucleotides used in this study.

Name	Sequence (5'>3') *
oDB0071	CTGAGCTCCTGCAGTAAAGGAGAAAATTTTATGTTTTCTAAAAAAAATGAT
oDB0072	TAGGATCCGGTTAATTAATTTTCTCTACAGCT
oCDTcdCmyc3	TAGGATCCTTATAAATCTTCTTCACTTATTAATTTTTGTTCTAAATCTTCTTCACTTATTAATTTT
oCD_SortaseF	GTCTGAGCTCCTGCAGTAAAGGAGAAAATTTTATGTTGAAAAAATTATATAGAATAG
oCD_SortaseR	CCCTCGAGAAATCAATCTACCATGAATCAC
oTcdCRev	AAACTCGAGAATTAATTTTCTCTACAGCTATCCCTGG
CDTcdCC51SF	CAATATATCCTCACCAGCTAGTTCTGAAGACCATGAGGAG
CDTcdCC51SR	CTCCTCATGGTCTTCAGAACTAGCTGGTGAGGATATATTG
oJC424	CAATATATCCTCACCAGCTGCTTCTGAAGACCATGAGGAG
oJC425	CTCCTCATGGTCTTCAGAAGCAGCTGGTGAGGATATATTG

* Restriction enzyme sites used underlined

Construction of the *tcdC-3xmyc* for cysteine accessibility analysis

To construct the expression constructs for the *tcdC-3xmyc*, the *tcdC* gene (CD0664 from *C. difficile* 630; GenBank accession no. YP_001087138.1) was amplified by PCR using primers oDB0071 and oDB0072 from *C. difficile* chromosomal DNA. The PCR product was subsequently cloned into pCR2.1TOPO (Invitrogen), according to manufacturer's instructions, to yield vector pCR2.1TcdC (Table 1). The TcdC fragment was amplified from pCR2.1TcdC with primers oDB0071 and oCDTcdCmyc3, which allows the addition of a 3xmyc-tag at the C-terminus. The resulting PCR fragment was digested with *SacI* and *BamHI* and ligated into pRPF185 digested with the same enzymes, to yield vector pLDJ1 (Table 1), placing *tcdC* under the control of the anhydrotetracycline inducible promoter (P_{tet}).

To mutate the cysteine at position 51 in TcdC, we used oligos CDTcdCC51SF and CDTcdCC51SR for the C51S mutation and the oligos oJC424 and oJC425 for the mutation C51A, respectively, in a Quikchange reaction, using pCR2.1-TcdC as a template. The mutated TcdC sequences were subsequently subcloned into pRPF185 as described for wild type *tcdC* above, yielding plasmids pLDJ2 and pJC129.

Construction of the HiBiT^{opt} fusions

The *hupA* gene (CD3496 from *C. difficile* 630; GenBank accession no. NC_009089.1) fused at the 3' end to the *hiBiT^{opt}* codon-optimized coding sequence (HupA-HiBiT^{opt}) was synthesized by Integrated DNA Technologies, Inc. (IDT). The synthesized fragment (full sequence in supplemental information) was digested with *BamHI* and *SacI*, and cloned into pRPF185 digested with same enzymes, yielding vector pAF302 (Table 1), which can be requested through Addgene (Addgene ID 137752).

The *srtB gene* (CD2718 from *C. difficile* 630; GenBank accession no. YP_001089230.1) was amplified from *C. difficile* genomic DNA with primers oCD-sortaseF/oCD-sortaseR, digested with *XhoI* and *SacI*, and placed into similarly digested pAF302, yielding vector pAP233 (Table 1) which can be requested through Addgene (Addgene ID 137753).

The *tcdC* gene was amplified by PCR with primer set oDB0071/oTcdCRev from *C. difficile* genomic DNA). The PCR fragment was digested with *XhoI* and *SacI* and cloned into similarly digested pAF302, yielding vector pJC111 (Table 1). To generate the equivalent C51S construct, *tcdC-C51S* was amplified with oBH0071 and oTcdCRev from oLDJ2, and cloned after *SacI-XhoI* digestion into similarly digested pCD111 to yield pJC127.

Affinity purification of the anti-TcdC polyclonal antibodies

Polyclonal antibodies against TcdC were raised in rabbits using the peptide CQLARTPDDYKYKKV [17]. To reduce background signals, serum from the final bleed from the immunized rabbits was subjected to affinity purification. Recombinant soluble 10xHis-TcdCΔN50 (lacking the N-terminal 50 amino acids) [23] was blotted on a PVDF membrane. The blot was stained with Ponceau S solution (0,2% (w/v) Ponceau S, 1% acetic acid) for 5 minutes. Subsequently, the blot was washed with TBST (500 mM NaCl, 20 mM Tris, pH 7,4, 0,05% (v/v) Tween-20) until the TcdC band was visible. The band was cut out of the blot and the piece of membrane was washed with Ponceau destaining solution (PBS, 0,1% NaOH) for 1 minute. Subsequently, the blot piece was washed twice with TBST for 5 minutes. The membrane was soaked in acidic glycine buffer (100 mM glycine, pH 2,5) for 5 minutes and washed twice in TBST for 5 minutes. The blot was blocked in 10% ELK (in TBST) for 1 hour at room temperature. After washing it twice with TBST for 5 minutes, the blot was incubated with 10 ml of 5 x diluted serum (in TBS) overnight at 4 °C. Afterwards, the diluted serum was removed and the blot was washed 3 times with TBST for 5 minutes and 2 times with PBS for 5 minutes. The blot was incubated with 1 mL of acidic glycine buffer for 10 minutes to elute the antibodies. The eluted antibody solution was immediately neutralized by adding 1M Tris-HCl, pH 8.0. After addition

of sodium azide (5 mM) and BSA (1 mg/ml), the affinity purified antibodies were stored at 4°C until use in experiments.

Cysteine Accessibility analysis

To perform the Cysteine Accessibility analysis, 25 ml *C. difficile* cultures were induced with 200 ng/ml anhydrotetracycline (ATc) at an OD_{600} of 0.3, for 2 hours. The culture was centrifuged for 10 min at 2800 x g and the pellet was frozen at -20°C until needed. Subsequently, the pellet was resuspended in 600 µl GTE buffer (50 mM glucose; 20 mM tris-HCl; 1 mM EDTA, pH7.5) supplemented with Complete protease inhibitor cocktail (CPIC, Roche Applied Science) and aliquoted in 100 µl. For the cysteine labelling, 1 mM of MPB was added and the samples were incubated for 15 min on ice. The reactions were quenched by the addition of 73 mM β-mercaptoethanol. Samples were washed twice in GTE buffer + CPIC and centrifuged at 2800 x g. Pellets were resuspended in 100 µl solubilization buffer (50 mm Tris-HCl, pH 8.1, 2% SDS, 1mM EDTA) with mixing for 5 min and sonication (2 × 5-10 sec pulses). To remove unspecific binding, 400 µl 0.2% PBS-T + Triton X-100 + CIPC was added to the sample together with 30 µl 50% protein A sepharose CL-4B (Amersham) slurry in phosphate buffered saline (PBS) supplemented with 1% bovine serum albumin (BSA), previously equilibrated in PBS/BSA 1%. After overnight incubation at 4 ˚C, protein A sepharose beads were removed by gentle centrifugation.

For immunoprecipitation, 50 µl of 50% protein A sepharose CL-4B (Amersham) slurry was added to each sample with the affinity purified polyclonal rabbit anti-TcdC antibody (1:200) and incubated at 4°C with gentle mixing for 2 hours. The slurry was pelleted (4000 × g) and washed 2 times with TENT buffer (150 mm NaCl; 5 Mm EDTA; 50 mm Tris; Triton-X-100 0.5%; pH 7.5) + 1% BSA + 0.5 M NaCl, 2 times with TENT + 1% BSA + 0.25M NaCl and 2 times with TENT buffer. The pellet was resuspended in 50 µl SDS loading buffer (250 mM Tris-Cl pH 6.8, 10 % SDS, 10% β-mercaptoethanol, 50% glycerol, 0.1% bromophenol blue) and incubated at 50°C for 5 min. Samples were spun down prior to SDS-PAGE analysis.

Immunoblotting and detection

Proteins were separated on a 12% SDS-PAGE gel and transferred onto polyvinyl difluoride (PVDF) membranes (Amersham), according to the manufacturers' instructions. The membranes were probed with antibodies monoclonal mouse anti-myc (α-myc, 1:1500, Invitrogen) or mouse anti-biotin (α-biotin, 1:1000) in PBST. After washing the blots with PBST, a secondary goat anti-mouse-HRP antibody (1:1000, Dako) was used. Bands were visualized

using the Clarity ECL Western Blotting Substrates (Bio-Rad) on an Alliance Q9 Advanced machine (Uvitec).

HiBiT^{opt} Assay

C. difficile cells were induced with 50 ng/mL ATc at an OD_{600} of 0.3-0.4, for 45 min. A one ml sample was collected and centrifuged (4000 x g) for SDS-PAGE analysis and for luminescent detection of HiBiT^{opt}-tagged proteins on a blot. 50 µL samples before and after centrifugation were collected for analysis.

To measure luciferase activity pellets where resuspended in 1mL PBS and 50 µL sample was taken for further luciferase detection. Samples were centrifuged (20000 x g) for 10 min and pellets incubated in 950 µL Lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.1 mg/mL lysozyme, CPIC) for 1h at 37°C. 50 µL sample was taken for further luciferase detection.

Samples were incubated with 50 µL Nano-Glo HiBiT Extracellular Detection System, a mixture of the NanoLuc LgBiT protein and luciferase substrate in buffer (Promega) for 5 min, in a 96-well white F-bottom plate. Luciferase activity was measured on a GloMax Multi+ instrument (Promega), with a 0.5 s integration time. All luciferase measurements were taken immediately after sampling. Data was normalized to OD_{600} of the culture the samples were derived from and statistical analysis was performed with Prism 7 (GraphPad Inc, La Jolla, CA) by two-way ANOVA.

For luminescent detection of HiBiT^{opt}-tagged proteins on a blot, total protein was resolved on a 12% SDS-PAGE gel and transferred onto polyvinyl difluoride (PVDF) membranes (Amersham). The membranes were washed with TBST and incubated with 200-fold diluted LgBiT protein in TBST (N2421, Promega), for 1 hour at room temperature. 500-fold diluted Nano-Glo Luciferase Assay Substrate (N2421, Promega) was added and incubated for 5 min at room temperature with gentle shaking. The membranes were analysed using an Alliance Q9 Advanced machine (Uvitec).

Images were prepared for publication in CorelDRAW Graphics Suite X7 software.

Acknowledgments

Work in the group of WKS is supported by a Vidi Fellowship (864.10.003) of the Netherlands Organization for Scientific Research (NWO) and a Gisela Thier Fellowship from the Leiden University Medical Center.

Supplemental Information

>*hupA-hiBiT*^{opt}

GAGCTCCTGCAGTAAAGGAGAAAATTTTGTGAATAAAGCTGAATTAGTATCAAAGATGGCAGAAAA
AAGTGGATTAACAAAGAAGGAAGCAGAAGCTGCGTTAAACGCATTTATGAGTTCTGTTCAAGATGC
ACTAGTAAATAATGAAAAAGTTCAATTAGTTGGATTTGGAACATTTGAGACAAGAGAAAGAGCTGCT
AGACAAGGAAGAAATCCAAGAGATCCAGAGCAAGTTATAGATATACCAGCTTCTAAAGCACCAGTTT
TCAAAGCTGGAAAAGGATTAAAGGATATAATAAATGGATCTCGAGGGGGTTCTAGTGGTGGTGGTG
GTTCTGGTGGTGGTGGTTCTAGTGGTGTTAGTGGTTGGAGACTTTTTAAGAAAATTTCTTAGGGATC
C

Supplemental Figure

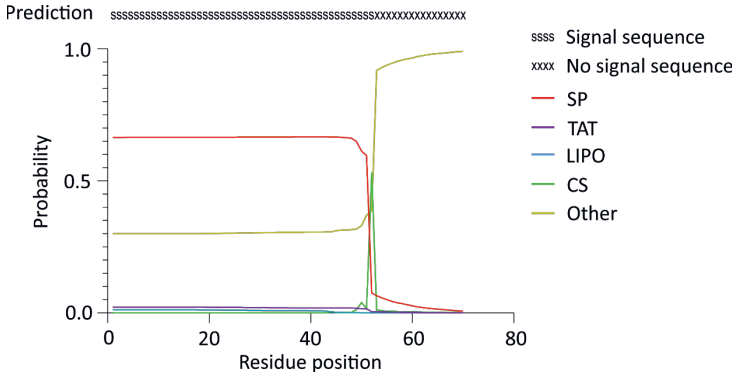


Fig. S1 - SignalP5.0 prediction for TcdC-C51S. Output from the SignalP 5.0 [26] web-server for the TcdC-C51S amino acid sequence. No signal sequence was detected (X). Probabilities of signal peptides presence from the systems Sec (SP, red line), Tat (purple line), and lipoprotein (LIPO, blue line) are shown. Predicted cleavage site score (CS, green line) and no signal sequence probability is depicted (OTHER, light green line).

References

1 Lawson, P. A., Citron, D. M., Tyrrell, K. L. & Finegold, S. M. Reclassification of *Clostridium difficile* as Clostridioides difficile (Hall and O'Toole 1935) Prevot 1938. *Anaerobe* **40**, 95-99 (2016).

2 Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H. & Kuijper, E. J. *Clostridium difficile* infection. *Nature Reviews Disease Primers* **2**, 16020 (2016).

3 Czepiel, J. *et al. Clostridium difficile* infection: review. *Eur J Clin Microbiol Infect Dis* **38**, 1211-1221 (2019).

4 Schaffler, H. & Breitruck, A. *Clostridium difficile* - From Colonization to Infection. *Front Microbiol* **9**, 646 (2018).

5 Chandrasekaran, R. & Lacy, D. B. The role of toxins in *Clostridium difficile* infection. *FEMS microbiology reviews* **41**, 723-750 (2019).

6 Rupnik, M. *et al.* Revised nomenclature of *Clostridium difficile* toxins and associated genes. (2005).

7 Dingle, K. E. *et al.* Evolutionary history of the *Clostridium difficile* pathogenicity locus. *Genome Biol Evol* **6**, 36-52 (2014).

8 Aktories, K., Papatheodorou, P. & Schwan, C. Binary *Clostridium difficile* toxin (CDT) - A virulence factor disturbing the cytoskeleton. *Anaerobe* **53**, 21-29 (2018).

9 Olling, A. *et al.* Release of TcdA and TcdB from *Clostridium difficile* cdi 630 is not affected by functional inactivation of the tcdE gene. *Microbial pathogenesis* **52**, 92-100 (2012).

10 Matamouros, S., England, P. & Dupuy, B. *Clostridium difficile* toxin expression is inhibited by the novel regulator TcdC. *Molecular microbiology* **64**, 1274-1288 (2007).

11 Dupuy, B. & Sonenshein, A. L. Regulated transcription of *Clostridium difficile* toxin genes. *Molecular microbiology* **27**, 107-120 (1998).

12 Karlsson, S. *et al.* Expression of *Clostridium difficile* toxins A and B and their sigma factor TcdD is controlled by temperature. *Infection and immunity* **71**, 1784-1793 (2003).

13 Dineen, S. S., Villapakkam, A. C., Nordman, J. T. & Sonenshein, A. L. Repression of *Clostridium difficile* toxin gene expression by CodY. *Molecular microbiology* **66**, 206-219 (2007).

14 Karlsson, S., Burman, L. G. & Akerlund, T. Induction of toxins in *Clostridium difficile* is associated with dramatic changes of its metabolism. *Microbiology* **154**, 3430-3436 (2008).

15 Hundsberger, T. *et al.* Transcription analysis of the genes tcdA-E of the pathogenicity locus of *Clostridium difficile*. *Eur J Biochem* **244**, 735-742 (1997).

16 Govind, R., Vediyappan, G., Rolfe, R. D. & Fralick, J. A. Evidence that *Clostridium difficile* TcdC is a membrane-associated protein. *J Bacteriol* **188**, 3716-3720 (2006).

17 Bakker, D., Smits, W. K., Kuijper, E. J. & Corver, J. TcdC does not significantly repress toxin expression in *Clostridium difficile* 630DeltaErm. *PloS one* **7**, e43247 (2012).

18 Merrigan, M. *et al.* Human hypervirulent *Clostridium difficile* strains exhibit increased sporulation as well as robust toxin production. *Journal of Bacteriology* **192**, 4904-4911 (2010).

19 Warny, M. *et al.* Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. *Lancet* **366**, 1079-1084 (2005).

20 Carter, G. P. *et al.* The anti-sigma factor TcdC modulates hypervirulence in an epidemic BI/NAP1/027 clinical isolate of *Clostridium difficile*. *PLoS pathogens* **7**, e1002317 (2011).

21 Murray, R., Boyd, D., Levett, P. N., Mulvey, M. R. & Alfa, M. J. Truncation in the tcdC region of the *Clostridium difficile* PathLoc of clinical isolates does not predict increased biological activity of Toxin B or Toxin A. *BMC Infectious Diseases* **9**, 103 (2009).

22 Cartman, S. T., Kelly, M. L., Heeg, D., Heap, J. T. & Minton, N. P. Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the tcdC genotype and toxin production. *Appl Environ Microbiol* **78**, 4683-4690 (2012).

23 van Leeuwen, H. C., Bakker, D., Steindel, P., Kuijper, E. J. & Corver, J. *Clostridium difficile* TcdC protein binds four-stranded G-quadruplex structures. *Nucleic acids research* **41**, 2382-2393 (2013).

24 Sonnhammer, E. L., von Heijne, G. & Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**, 175-182 (1998).

25 Tsirigos, K. D., Peters, C., Shu, N., Kall, L. & Elofsson, A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic acids research* **43**, W401-407 (2015).

26 Armenteros, J. J. A. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature biotechnology* **37**, 420 (2019).

27 Bogdanov, M., Zhang, W., Xie, J. & Dowhan, W. Transmembrane protein topology mapping by the substituted cysteine accessibility method (SCAM(TM)): application to lipid-specific membrane protein topogenesis. *Methods* **36**, 148-171 (2005).

28 Fagan, R. P. & Fairweather, N. F. *Clostridium difficile* has two parallel and essential Sec secretion systems. *J Biol Chem* **286**, 27483-27493 (2011).

29 Oliveira Paiva, A. M., Friggen, A. H., Hossein-Javaheri, S. & Smits, W. K. The Signal Sequence of the Abundant Extracellular Metalloprotease PPEP-1 Can Be Used to Secrete Synthetic Reporter Proteins in *Clostridium difficile*. *ACS synthetic biology* (2016).

30 Oliveira Paiva, A. M. *et al.* The Bacterial Chromatin Protein HupA Can Remodel DNA and Associates with the Nucleoid in *Clostridium difficile*. *Journal of molecular biology* **431**, 653-672 (2019).

31 Eggers, C. *et al.* Monitoring Protein Dynamics at Endogenous Levels with a Luminescent Peptide Tag. (Promega Corporation, USA, 2018).

32 Schwinn, M. K. *et al.* CRISPR-Mediated Tagging of Endogenous Proteins with a Luminescent Peptide. (2017).

33 Jacobitz, A. W., Kattke, M. D., Wereszczynski, J. & Clubb, R. T. Sortase Transpeptidases: Structural Biology and Catalytic Mechanism. *Adv Protein Chem Struct Biol* **109**, 223-264 (2017).

34 van Eijk, E. *et al.* Complete genome sequence of the *Clostridium difficile* laboratory strain 630Deltaerm reveals differences from strain 630, including translocation of the mobile element CTn5. *BMC Genomics* **16**, 31 (2015).

35 Chambers, C. J., Roberts, A. K., Shone, C. C. & Acharya, K. R. Structure and function of a *Clostridium difficile* sortase enzyme. *Scientific reports* **5**, 9449-9449 (2015).

36 Schaffer, C. & Messner, P. Emerging facets of prokaryotic glycosylation. *FEMS microbiology reviews* **41**, 49-91 (2017).

37 Chekli, Y. *et al.* Visualizing the dynamics of exported bacterial proteins with the chemogenetic fluorescent reporter FAST. 2020.2005.2012.090142 (2020).

38 Streett, H. E., Kalis, K. M. & Papoutsakis, E. T. A Strongly Fluorescing Anaerobic Reporter and Protein-Tagging System for *Clostridium* Organisms Based on the Fluorescence-Activating and Absorption-Shifting Tag Protein (FAST). *Appl Environ Microbiol* **85** (2019).

39 Ginalski, K., Kinch, L., Rychlewski, L. & Grishin, N. V. BOF: a novel family of bacterial OB-fold proteins. *FEBS Lett* **567**, 297-301 (2004).

40 Arcus, V. OB-fold domains: a snapshot of the evolution of sequence, structure and function. *Curr Opin Struct Biol* **12**, 794-801 (2002).

41 Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* **382**, 769-781 (2004).

42 Ethapa, T. *et al.* Multiple factors modulate biofilm formation by the anaerobic pathogen *Clostridium difficile*. *J Bacteriol* **195**, 545-555 (2013).

43 Dubois, T. *et al.* A microbiota-generated bile salt induces biofilm formation in *Clostridium difficile*. *NPJ biofilms and microbiomes* **5**, 14 (2019).

44 Patel, D., Wines, B. D., Langley, R. J. & Fraser, J. D. Specificity of staphylococcal superantigen-like protein 10 toward the human IgG1 Fc domain. *Journal of immunology* **184**, 6283-6292 (2010).

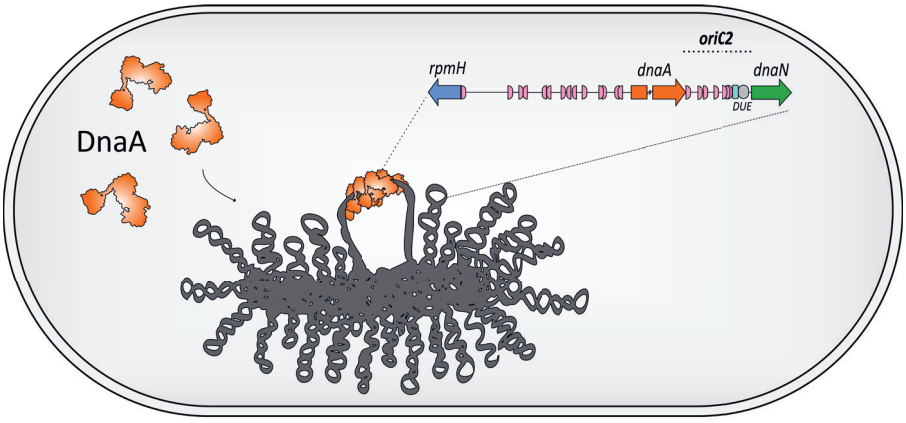
45 Pilonieta, M. C., Erickson, K. D., Ernst, R. K. & Detweiler, C. S. A protein important for antimicrobial peptide resistance, YdeI/OmdA, is in the periplasm and interacts with OmpD/NmpC. *J Bacteriol* **191**, 7243-7252 (2009).

46 Moreira, C. G. *et al.* Virulence and stress-related periplasmic protein (VisP) in bacterial/host associations. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 1470-1475 (2013).

47 Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular cloning : a laboratory manual*. (Cold Spring Harbor Laboratory, 1989).

48 Purdy, D. *et al.* Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Molecular microbiology* **46**, 439-452 (2002).

49 Hussain, H. A., Roberts, A. P. & Mullany, P. Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630Δerm) and demonstration that the conjugative transposon Tn916DeltaE enters the genome of this strain at multiple sites. *J Med Microbiol* **54**, 137-141 (2005).



Clostridioides difficile