7

Clostridium difficile TcdC protein binds four stranded G- quadruplex structures

H C. van Leeuwen[1], Dennis Bakker[1], Philip Steindel[2], Ed J. Kuijper[1], Jeroen Corver[1]

[1] *Department of Medical Microbiology, Center of Infectious Diseases, Leiden University Medical Center, the Netherlands.*
[2] *Department of Biochemistry, Brandeis University, MS009, 415 South Street, Waltham, MA 02454*

Nucleic Acids Res 2013

Abstract

Clostridium difficile infections are increasing worldwide due to emergence of virulent strains. Infections can result in diarrhea and potentially fatal pseudomembranous colitis. The main virulence factors of *C.difficile* are clostridial toxins TcdA and TcdB. Transcription of the toxins is positively regulated by the sigma factor TcdR. Negative regulation is believed to occur through TcdC, a proposed anti-sigma factor. Here, we describe the biochemical properties of TcdC in order to understand the mechanism of TcdC action.

Bioinformatic analysis of the TcdC protein sequence predicted the presence of a hydrophobic stretch (amino-acids 30-50), a potential dimerization domain (aa 90-130) and a C-terminal OB-fold (oligonucleotide-binding).

Gel filtration chromatography of two truncated recombinant TcdC proteins (TcdCΔ1-89 and TcdCΔ1-130), showed that the domain between aa 90 and 130 is involved in dimerization.

Binding of recombinant TcdC to single-stranded DNA was studied using a single-stranded SELEX approach. This involved specific binding of ss-DNA sequences from a pool of random oligonucleotides, as monitored by electrophoretic-mobility shift assay. Analysis of the oligonucleotides bound showed that the OB-fold domain of TcdC can bind specifically to DNA folded into G-quadruplex structures containing repetitive guanine nucleotides forming a four-stranded structure.

In summary, we provide evidence for DNA binding of TcdC, which suggests an alternative function for this proposed anti-sigma factor.

Introduction

Clostridium difficile is a spore forming, anaerobic bacterium that can cause antibiotic associated diarrheal disease in humans. In the past decade, the incidence, complications and mortality of *Clostridium difficile*-associated infection (CDI) have increased dramatically due to the emergence of new hypervirulent PCR Ribotypes (RT) (1-4). Virulence of *C.difficile* has been linked to the production of two toxin molecules, Toxin A and Toxin B, which are encoded within the Pathogenicity Locus (PaLoc). These toxins cause intestinal damage and ultimately clinical disease (5). Both toxins have the same enzymatic activity. Upon entering intestinal epithelial cells they catalyze the transfer of glucose onto the Rho family of GTPases, leading to reorganization of the actin cytoskeleton, complete rounding of cells and destruction of the intestinal barrier function. This causes diarrhea and in some cases may lead to a severe inflammatory response and pseudomembranous colitis.

The mechanisms that regulate the levels of toxin synthesis are slowly being unraveled. Toxin genes, *tcdA* and *tcdB*, are located on the PaLoc together with two regulatory genes *tcdR* and *tcdC*, and *tcdE*, which encodes a holin like protein that may facilitate the release of the toxins into the extracellular environment (6). TcdR has been demonstrated to activate gene expression of both toxins as a specific RNA polymerase sigma factor belonging to the subgroup of extracytoplasmic function (ECF) σ^{70}-family of RNA polymerase sigma factors (7). Members of this group include several σ-factors involved in positive regulation of potent toxins such as botulinum neurotoxin (BotR of *Clostridium botulinum*) and tetanus neurotoxin (TetR of *Clostridium tetani*) (8). Toxin expression is also influenced by the nutritional status of the bacteria; a rapidly metabolizable carbon source such as glucose inhibits toxin expression (9). In addition, general regulatory molecules, such as CodY and CcpA are known to influence toxin synthesis (10;11). In *C.difficile*, TcdR not only stimulates toxin gene transcription but also activates its own expression suggesting a large overshoot in protein expression once activated (7). A negatively acting mechanism therefore is required to put a limit on this system during unrestricted growth of *C.difficile*.

Activation of bacterial gene expression by specific σ-factors is often subject to control by specific antagonists, called anti-sigma factors (12). Generally they sequester their cognate σ factor, preventing it from interacting with the RNA polymerase. Encoded within the PaLoc is TcdC, which has been postulated to act as an anti-sigma factor and negatively regulate toxin production. TcdC transcription pattern was reported to be inverse to TcdR and the toxins as it is highly transcribed and expressed during the exponential growth phase, while its expression is strongly reduced as the growth rate slows in stationary phase (13). This inverse correlation

suggested that TcdC interferes with toxin gene expression. However, more recent studies have shown that this inverse correlation cannot be confirmed using quantitative RT-PCR (14-16). This suggests that TcdC may not be as important in toxin regulation as previously thought.

A direct inhibitory effect on transcription of *tcdC* has been shown *in vitro*. The TcdR-RNA-polymerase-DNA complex is destabilized by TcdC, preventing initiation of transcription. However, once a stable open complex is formed with the promoter no inhibition by TcdC occurs (17). The target of TcdC in (prevention of) complex formation is unclear; no interaction with the TcdR-RNAP-complex was found nor does TcdC bind to dsDNA in the promoter, suggesting a potentially unique inhibitory mechanism of TcdC.

Recent *in vivo* studies on the importance of TcdC on toxin expression show contradictory results. *TcdC* complementation of strain M7404, a toxinogenic strain that lacks a functional *tcdC* gene, results in a reduced amount of produced toxin and an attenuated phenotype in hamsters (18). In contrast, complementation of strain R20291, another strain that lacks a functional *tcdC* gene, with a functional *tcdC* gene did not alter the toxin titers (19). In addition, knockout of *tcdC* in strain 630Δ*erm* did not result in an increased level of toxins produced, nor did it result in increased toxin mRNA production (14).

Since the suggested anti sigma function of TcdC is not undisputed and because the mechanism by which TcdC is supposed to inhibit TcdR mediated transcription is unknown, we aimed to further characterize the biochemical properties of TcdC. Through *in silico* analyses we found that TcdC contains a predicted single-stranded nucleic acid binding fold (OB-fold). In this paper we show for the first time through a combination of *in silico* analysis and biochemical experiments that TcdC can bind to nucleic acids.

Material and Methods

Construction of plasmids

To construct his10-tagged TcdC expression plasmids, the sequence was amplified by PCR from *C.difficile* strain 630 genomic DNA, using specific primers, see Table 1. The PCR products were digested with NdeI and XhoI or NdeI and BamHI and ligated into pET16b (Novagen) similarly digested with NdeI and XhoI/BamHI. This resulted in the construction of TcdC expression vectors containing a 10-His-tag at its N-terminus.

Table 1: Primers used to generate bacterial expression constructs.

Delta1-50 TcdC
Forward primer, TATGCATATGGGATATGATACTGGTATTAC Reverse primer, TTTTCTCGAGTTAATTAATTTTCTCTACAGCTATCCC
Delta1-89 TcdC
Forward primer, GTTCCATATGAAAGACGACGAAAAGAAAGCTATTG Reverse primer as for Delta1-50 TcdC
Delta1-130 TcdC
Forward primer, TATGCATATGGGATATGATACTGGTATTAC Reverse primer as for as for Delta1-50 TcdC
Delta Delta1-89; Delta208-232 TcdC (90-207)
Forward primer as for Delta1-89 TcdC Reverse primer, TACTGGATCCTTTAAGCACTTATACCTCTTATAG

DNA binding studies

Probes used for band shift assays were obtained from Eurogentec (Maastricht, The Netherlands), end labeled with T4-polynucleotide kinase and ^{32}P-γ-ATP and purified using Micro Bio-Spin Colums P-30 Tris RNase Free (Biorad) according to manufacturer's instructions. Binding reactions were carried out for 60 minutes on ice in 20 ul binding buffer (20 mM Hepes-KOH pH 7.5, 50 mM NaCl, 40 mM KCl, 7% glycerol, 1 mM EDTA, 0.1 mM DTT and 0.25 pmol probe (12.5 nM). Free DNA and protein-DNA complexes were separated on a 7% polyacrylamide gel (37.5:1) run in 0.5xTBE. Dried gels were exposed to a Biorad phosphoimaging screen-K and scanned on a Typhoon 9410 from GE Healthcare. The equilibrium dissociation constant (Kd) was calculated at half saturation Kd=Pt-Db (Db=DNA bound, 6 nM). The Pt (total protein concentration) was calculated using a deduced molecular mass of 18.8 kD for the His10-TcdCΔ1-89 protein.

Single stranded SELEX

The random site used in the first selection round was 5' A G T G C A G T G G A T C C T G T C G - N N N N N N N N N N N N N N N - AGGCGAATTCAGTCCAAGTG 3', ^{32}P-labelled at the 5'end. Binding reactions were as described above with 10 or 100 ng purified TcdC Δ1-89. In the first selection round 10 ul of 50% Cobalt^{2+}-beads (Clontech) in binding buffer was added to the binding reaction and incubated for 30 min at 4°C under continuous rotation. Subsequently the beads were spun down (30 sec. at 100 g) and washed three times in 100 ul binding buffer. Bound random oligo was eluted in 20 ul binding buffer with 250 mM Imidazol and subsequent heating heating 10 minutes at 95 °°C. Beads were spun down and supernatant collected. 5 ul of supernatant was amplified

using 30 PCR cycles with primers, AGTCAGTGCAGTGGATCCTGTCG (forward) and CACTTGGACTGAATTCGCCTC (reverse). The resulting 53 bp PCR product was digested with N.BtsI nicking endonuclease (New England Biolabs) and labeled using ^{32}p-γ-ATP. Digested and labeled probe was separated on a 12% polyacrylamide gel (19:1). Following exposure of the wet gel to X-ray film (Fujifilm, super RX), the highest band corresponding to the uncleaved top stand was cut out and eluted according to the QIAEX II Gel Extraction Kit for polyacrylamide Gels (Qiagen). 5 ul of the extracted probe was used in a next round of selection using bandshift assay. The Protein-DNA complexes were separated on a 7% polyacrylamide gel as described above. After exposure to film the bound probe (shift) was cut out and eluted for an additional round of selection.

Quadruplex staining

Specific staining of quadruplex forming DNA was performed according to Yang *et al.* (20). Briefly, polyacrylamide gels were incubated in 20 uM ETC ($C_{39}H_{47}N_3O_6S_4$, Organica Feinchemie GmbH Wolfen) in phosphate buffer saline (PBS) for 30 minutes. Rinsed five times with water and then scanned in a Typhoon 9410 (GE Healthcare), excitation 532 nm and emission 610 nm.

Bioinformatics analysis

For all bioinformatic analyses protein sequence Q189K7 (*C.difficile* strain 630) was used. Predictions of coiled-coil helices were carried out using the Multicoil Scoring Form ((21); http://groups.csail.mit.edu/cb/multicoil/cgi-bin/multicoil.cgi). All predictions were performed using standard settings. All sequence alignments were performed by use of Clustal Omega - Multiple Sequence Alignment, available from EMBL-EBI European bioinformatics institute (http://www.ebi.ac.uk/Tools/msa/clustalo/; Sievers *et al.*, Mol Syst Biol. 2011).

Structural models of the TcdC conserved C-terminal domain were generated by the automated I-TASSER (threading, assembly and refinement) simulation method; http://zhanglab.ccmb.med.umich.edu/I-TASSER/ (22;23). Predictions were done using the standard parameters. As part of sequence homology detection, the protein alignment was analyzed using HHpred at the Max-Planck Institute for Developmental Biology (http://toolkit.tuebingen.mpg.de/hhpred; (24)). Predictions were done using the following parameters: Selected, database pfamA_v26.0; Max. MSA Generation iterations, 0. Other parameters set at default.

Protein purification

E.coli (BL21) lysates (50 mM sodium phosphate buffer, pH 8.0, 5 mM beta-mercaptoethanol, 0.1% NP40, 300 mM NaCl) containing histidine tagged proteins

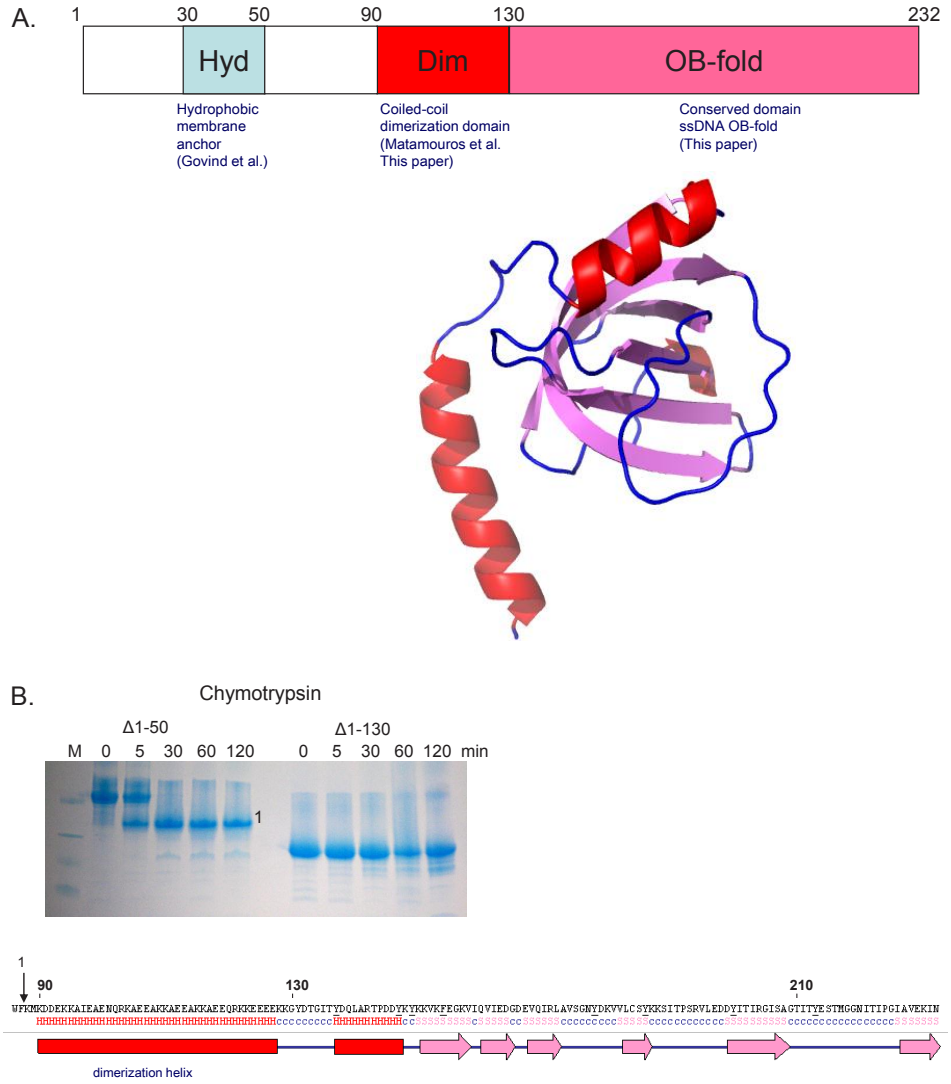

Figure 1: Domain organization of *Clostridium difficile* TcdC. **(A)**. Schematic presentation of TcdC with three predicted domains; Hyd= hydrophobic membrane anchor, Dim:dimerization domain and OB fold = conserved C-terminal domain containing a predicted oligonucleotide binding fold. Bottom, spatial structure of the TcdC- C-terminal domain (aa 90-232) predicted by the automated simulation method I-TASSER (Roy *et al.*, 2010) and represented with a ribbon diagram. N and C indicate amino-terminus and carboxy-terminus. **(B)** Limited proteolysis of TcdC⊿1-50 and TcdC⊿1-130. Histidine-tagged proteins were digested with chymotrypsin for 5, 30, 60, or 120 min, and the fragments were resolved by SDS-PAGE. Amino-terminal sequences were identified using an Edman degradation. Site of chymotrypsin cleavage (arrow 1) indicated in the primary amino acids sequence and TcdC predicted secondary structure elements. H=helix (red box), C=coil (blue line), S=sheet (pink arrow).

were loaded on a 1 mL Ni-NTA column (Qiagen). The column was washed with 20 mL wash buffer (50 mM sodium phosphate buffer pH 7.0, 300 mM NaCl, 5 mM mercaptoethanol, 5% glycerol, 20 mM Imidazol). The His-tagged proteins eluted at ~200 mM imidazole when employing a 25 mL linear gradient ranging from 20 to 250 mM imidazole. Peak fractions containing the His-tagged proteins were pooled and 200ul loaded onto a superdex 75 gelfiltration column equilibrated and run in 50 mM sodium phosphate buffer pH 7.0, 150 mM NaCl, 5 mM mercaptoethanol, 5% glycerol. Protein concentrations were measured and peak fractions were used for DNA binding studies.

Results

In silico analysis of TcdC

In many cases sequence similarity allows the inference of protein function. At the primary amino acid sequence level, the C-terminal domain of TcdC (residues 130-232, conserved domain, Figure 1A) has sequence identity (conservation) to potential/putative protein homologues from both anaerobic and facultative aerobic members of the Firmicutes phylum (Figure 2). Though several TcdC homologues have been identified, none of them have been characterized biochemically in detail and therefore do not provide a clue to the TcdC mechanism of action.

As the primary sequence gave no indication to its function we used computational protein structure prediction for detecting remote homologous templates. Structural models of the TcdC conserved C-terminal domain were generated by the automated I-TASSER (threading, assembly and refinement) simulation method (22;23). The best model (Figure 1A) was predicted to be composed of a five-stranded closed beta-barrel connected by large loops (C-score = -0.9; C-score is a confidence score for the predicted model with a C-score score above cutoff of − 1.5 is used to select models of correct topology). Matching the best predicted model with proteins from the PDB databasesrevealed a nucleic acid binding OB-fold (oligonucleotide binding, IPR016027) containing domain in all the 10 top matches/best scoring templates (TM = 0.8 – 0.7; TM-score >0.5 indicates a model of correct match topology). The core of the OB-fold forms a surface to bind to single-stranded DNA (ssDNA) or RNA (25;26). Variations in folds, loops and amino-acids in the binding interface determine ligand and sequence specificity. Members of this OB-fold group include proteins critical for DNA replication protein (RPA), DNA recombination (RuvA), translation (tRNA synthetase anticodon binding protein) and telomere-end-binding proteins (hPot-1)(25;26).

I-TASSER folding of the region preceding the conserved domain (aa 90-130) predicted a large helix (Figure 1A) containing many positive and negatively charged amino-acids. Such a helix clearly can form a charged coiled-coil motif with another molecule thereby forming an intertwined dimer as was predicted by Matamouros *et. al.* (17). The coiled-coil prediction was confirmed using the Multicoil Scoring Form (21), which calculated a maximum coiled-coil probability of 0.861 of this region (data not shown).

In addition to the protein structure prediction, we used the TcdC conserved domain protein alignment (Figure 2) for a highly sensitive profile-based search (24;27). Using the TcdC-conserved domain multiple sequence alignment rather than a single sequence as a query increases sensitivity and allows for homology detection of protein families. Pairwise comparison of the TcdC profile with the PFAM database resulted in a hit with PF12869, tRNA_anti-like family containing the nucleic acid-binding OB fold (E-value 1.3e-6, probability 98.3). In summary, these *in silico* analyses clearly suggest that TcdC forms a dimeric ssDNA binding OB-protein fold.

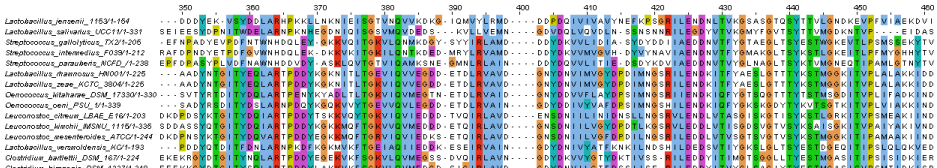

Figure 2: TcdC conserved domain protein alignment with protein homologues from members of the Firmicutes phylum.Numbering top according to alignment; Numbers bottom according to *Clostridium difficile* TcdC amino acids.

Limited proteolysis suggests a folded structure of the TcdC conserved domain

In order to confirm the predicted domains and borders of the nucleic acid binding OB-fold of TcdC, we cloned the *tcdC* gene including 10-histidine codons at the N-terminus into a bacterial expression vector. In order to produce soluble protein expression in *E.coli*, the first 50 amino-acids, which contain the reported hydrophobic membrane anchor (13) (Figure 1A) were removed. His-10 tagged TcdC was overexpressed in *E.coli* (BL21) and purified using a nickel affinity column.

To investigate the local conformation of this TcdC protein, we used limited proteolytic digestion. Protease resistance is an indication of structured protein sequences, as folded structure is usually protected from proteolytic degradation. TcdCΔ1-50 was digested with chymotrypsin, which cleaves after aromatic amino acids. Despite the presence of 11 potential cleavage sites (W, Y, F), chymotrypsin digestion of TcdCΔ1-50 led to only one distinct fragment (Figure 1B). To identify this fragment, the proteolytic product was subjected to N-terminal sequencing using Edman degradation. The identified N-terminal sequence (KMKD) corresponds to residue 88 of TcdC directly adjacent to the large coiled-coil helix. When we tested TcdCΔ1-130 (Figure 1B), corresponding to the OB-fold domain, we observe hardly any cleavage.

Taken together, these studies provide strong support for a folded structure of the TcdC conserved domain, including the dimerization domain, resistant to proteolytic cleavage.

TcdC contains a dimerization domain

Consistent with the proteolytic protection assay and prediction of the coiled-coil dimerisation helix, we constructed an expression vector containing the TcdC conserved domain including this putative dimerisation domain (TcdCaa90-232, here named TcdCΔ1-89). In addition a construct without the dimerization domain (TcdCΔ1-130) was generated. Both proteins were subsequently purified using nickel–affinity chromatography and gel-filtration (Figure 3).

Besides extra purity the latter column allows for separation by size and thus molecular weight estimation (Figure 4). Indeed the apparent molecular weight of TcdCΔ1-89 of 35 kDa, with a predicted molecular weight of 18 kDa, fits a dimeric protein. In contrast, TcdCΔ1-130, with a predicted molecular weight of 14 kDa and apparent molecular weight of 14 kDa, fits a monomeric protein. This confirms that the region between AA 90 – 130 contains a dimerization domain. Dimerization was confirmed using a cross linking with glutaraldehyde which can form stable intersubunit covalent bonds (Figure 4C). This experiment shows that TcdC Δ1-130 forms no visible dimers after crosslinking, while TcdC Δ1-89 forms

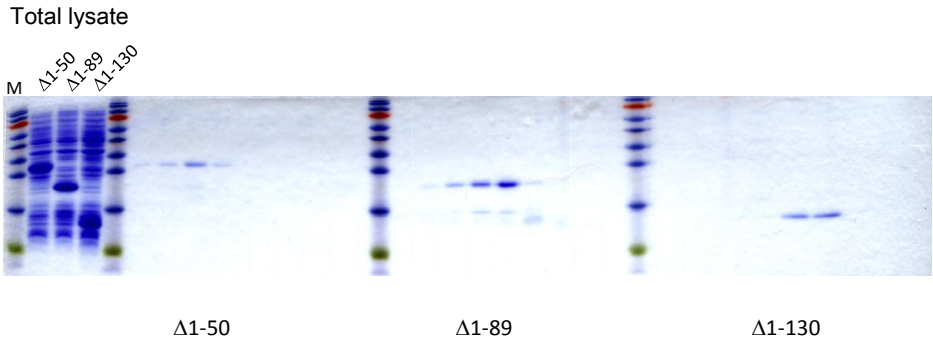

Figure 3. Cobalt2+ agarose and gel filtration purified His10-TcdC Δ1-50 , Δ1-89 and Δ1-130.

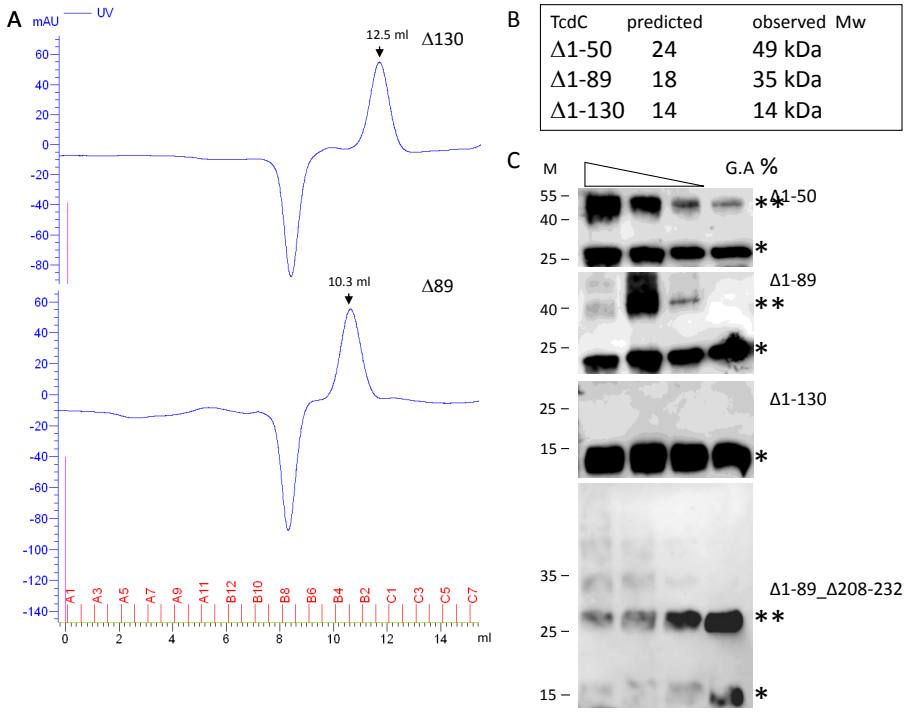

Figure 4: Gel filtration separation and cross linking of TcdC deletion mutants show that the TcdC dimerization domain is located between AA90-130. **(A)** Chromatographic separation of purified TcdC proteins on a superdex 75 column. Arrows indicate the elution points (mL) for the proteins. **(B)** Predicted molecular weight and observed molecular weight on the GF column of the respective deletion mutants. **(C)** TcdC dimerization as evidenced by glutaraldehyde cross-linking. 25 ng of TcdC was incubated for 30 minutes in the presence or absence of glutaraldehyde. Glutaraldehyde (G.A.) concentrations were 0.006% , 0.003%, 0.0006% and 0% (from left to right). After incubation, proteins were analyzed by Western blot analysis, using His-tag specific monoclonal antibody (Abcam). * indicates the bands corresponding to monomeric protein. ** corresponding to the dimer**.**

dimers already at low concentration of glutaraldehyde, thereby confirming the gel-filtration experiments.

TcdC does not bind to *tcdA* promoter elements

Based on existing evidence for the TcdC point of action i.e. destabilizing open complex formation before transcription initiation (17), we tested binding to (ss-) DNA corresponding to the region of the *tcdA* promoter that undergoes melting during transcription (opening of the double helix, resulting in exposed ssDNA) (28;29). Using protein TcdC Δ1-89 in a mobility shift assay, we tested binding to the *tcdA* double stranded promoter (-32 to +22 relative to transcription start), single stranded promoter top strand (non-template), single stranded promoter bottom strand (template), open promoter complex (region -10 to -6 or -13 to +4 open) as well as the *tcdA* mRNA gene transcript (+1 to +22) and the DNA-RNA hybrid (see Table 2). Surprisingly, we found no DNA binding for any of these fragments (data not shown). Also partially double stranded/single stranded overhang (5'and 3') and forked templates of the promoter (see Table 2) showed no binding. Finally, we tested a synthetic Holliday junction (30;31), which can be found at replication origins and recombination junctions, but found no binding.

TcdC binding sites selected through Selex

Because of our unsuccessful attempt to find the TcdC DNA binding site directly, we adapted a single-stranded SELEX (Systematic Evolution of Ligands by EXponential enrichment, (32), a procedure that allows extraction of oligomers with an optimal binding affinity from an initially random pool of oligonucleotides (Figure 5). After site-selection and PCR amplification, single stranded DNA is recreated using (asymmetric -) nicking of the bottom strand of the amplified selected sites followed by denaturation. These sites are subsequently used in an additional selection round thereby increasing the specificity of the selection procedure (Figure 6). Initial selection of his10-TcdCΔ1-89 bound fragments from the pool of ss-oligonucleotides, containing a stretch of 15 random nucleotides was performed through $Cobalt^{2+}$-agarose beads pull down (round 1, Figure 2). Two additional selection rounds were carried out using separation of bound DNA fragments on a polyacrylamide gel (round 2 and 3, Figure 6).

During these selection rounds we observed a higher molecular weight product (HMW) arise, which is bound and shifted by TcdCΔ1-89 (Figure 6). Each round showed a clear enrichment of the amount of higher molecular weight product being bound and shifted in the presence of TcdCΔ1-89. After three rounds of selection the enriched sites were cloned and sequenced. Table 3, shows the individual sites selected by the TcdCΔ1-89 ssSELEX. Most of the sequences selected (17/18)

Table 2: Primers used to test TcdC ss/ds DNA binding.

***tcdA* promoter double strand promoter** 795576..795629
-10 +1 5'CAAATTACTATCAGACAATCT**CCTTAT**CTAAT**A**GAAGAGTCAATTAACTAATTG 3' 3'GTTTAATGATAGTCTGTTAGAGGAATAGATTATCTTCTCAGTTAATTGATTAAC 5'
***tcdA* promoter template strand**
5'CAATTAGTTAATTGACTCTTCTATTAGATAAGGAGATTGTCTGATAGTAATTTG 3'
***tcdA* promoter non-template strand**
5'CAAATTACTATCAGACAATCTCCTTATCTAATAGAAGAGTCAATTAACTAATTG 3'
***tcdA* promoter large open promoter -13 to +4**
CTCCTTATCTAAT**A**GAA 5'CAAATTACTATCAGACAAT GAGTCAATTAACTAATTG 3' 3'GTTTAATGATAGTCTGTTA CTCAGTTAATTGATTAAC 5' CTCCTTATCTAATAGAA
***tcdA* promoter small open promoter -10 to -6**
TTAT 5'CAAATTACTATCAGACAATCT**CC** CTAAT**A**GAAGAGTCAATTAACTAATTG 3' 3'GTTTAATGATAGTCTGTTAGAGG GATTATCTTCTCAGTTAATTGATTAAC 5' TTAT
***tcdA* promoter 3'overhang**
5'CAAATTACTATCAGACAATCTCCTTATCTAAT**A**GAAGAGTCAATTAACTAATTG 3' 3'GTTTAATGATAGTCTGTTAGAGGAATAGATTA 5'
***tcdA* promoter 5'overhang**
5'CAAATTACTATCAGACAATCTCCTTATCTAAT**A**GAAGAGTCAATTAACTAATTG 3' 3' TCTTCTCAGTTAATTGATTAAC 5'
***tcdA* promoter vorked template**
TCTTCTCAGTTAATTGATTAAC 3' 5' CAAATTACTATCAGACAATCTCCTTATCTAAT 3'GTTTAATGATAGTCTGTTAGAGGAATAGATTA TCTTCTCAGTTAATTGATTAAC 5'
***tcdA* RNA transcript +1 to +22**
5'AGAAGAGUCAAUUAACUAAUUG 3'
***tcdA* DNA-RNA hybrid**
5'**A**GAAGAGUCAAUUAACUAAUUG 3' 3'GTTTAATGATAGTCTGTTAGAGGAATAGATTATCTTCTCAGTTAATTGATTAAC 5'
Four way junction
5'GACGCTGCCGAATTCTGGCTTGCTAGGACATCTTTGCCCACGTTGACCC 3' 5'TGGGTCAACGTGGGCAAAGATGTCCTAGCAATGTAATCGTCTATGACGTT 3' 5'CAACGTCATAGACGATTACATTGCTAGGACATGCTGTCTAGAGACTATCGA 3' 5'ATCGATAGTCTCTAGACAGCATGTCCTAGCAAGCCAGAATTCGGCAGCGT 3'

contain a stretch of 3 Gs (highlighted bold) and 2/3rd of the selected clones contains an A nucleotide preceding this G-stretch. 6 out of 18 clones contain two stretches of 3 Gs. Next, we selected two sites obtainedfrom the ssSELEX. One clone with a single aGGG consensus site (clone #2, Table 3) and one with a double consensus site (clone #5, Table 3) and tested these individually on a polyacrylamide gel. As shown in Figure 7A, both clones form the HMW product also observed in the ssSELEX, although in different efficiencies (on average ~90% of #5 formed HMW product compared to ~10% of clone #2). When we tested TcdC binding the HMW band was readily shifted, thereby confirming their efficient binding selection. In contrast to the HMW product, none of the ssDNA product was shifted (Figure 3A). Assuming a simple binding model, gel retardation experiments allow for a quick estimate of the protein equilibrium binding constants (Kd) at half saturation using the formula Kd=Pt-Db (the total protein and DNA concentrations at 50% binding). Quantification of the binding revealed a protein dissociation constant of ~30 nM (Figure 7B). When we assume that TcdC binds as a dimer, the Kd is ~15 nM. Typically, affinity of OB-fold proteins range from 1nM to 100 nM (33). Upon titration of TcdC, at least three complexes can be observed, which suggest multiple binding sites on the quadruplex structure (Figure 7B), with the first shifted complex representing the highest affinity binding site. Alternatively the dimeric protein might bind more than one quadruplex structure (each monomer bindings separate quadruplexes) thus forming larger complexes.

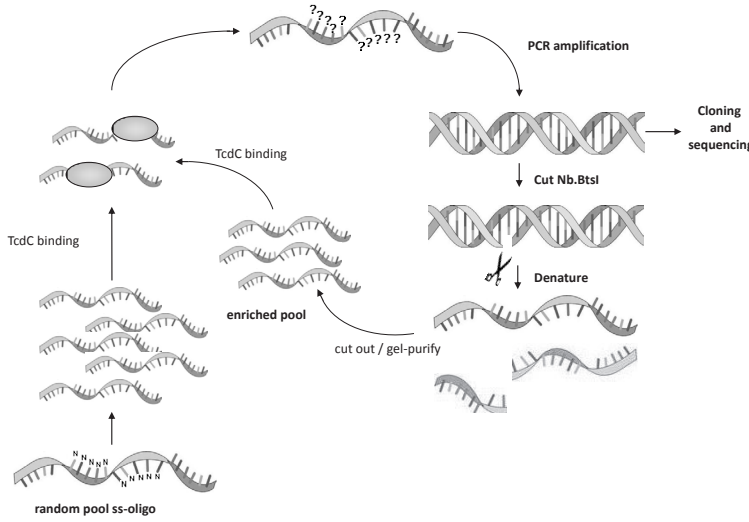

Figure 5: In vitro selection of TcdC binding sites (ssSELEX) starting with a random ssDNA oligonucleotide library, including generation of single stranded DNA after PCR amplification using a sequence specific nicking endonuclease (Nb.BtsI) that cleaves only one strand on a double stranded DNA product.

Characterization of the TcdC bound sequences

To characterize the structure of the HMW product and demonstrate that the HMW complex is the result of intra- or intermolecular structures, we heated radiolabeled clone #5 in the presence of formamide thereby denaturing DNA duplexes and secondary structures. Upon heating, the HMW product shifts to a lower molecular weight (Figure 8A), suggesting that the DNA element is a multiplex forming secondary structures. The fact that this unusual HMW structure is likely to be a multiplex and contains stretches of GGGs suggested that it could form a so called G- quadruplex, a four-stranded helical structure with four guanine bases from each strand forming hydrogen-bonds G-tetrads). Three (or more) guanine tetrads can stack on top of each other to form a G-quadruplex. We therefore tested the HMW product with ETC ($C_{39}H_{47}N_3O_6S_4$), an extended aromatic cyanine dye(that specifically recognizes stacked G-quadruplexes (20).

Figure 8B shows that ETC specifically stains the HMW product but not duplex and single-strand DNA (compare to Figure 8C), confirming that the selected element forms a G-quadruplex. To further support the quadruplex structure (QS) proposed for the #5 sequence we analyzed the HMW formation with a number of point mutations. We did not observe changes in the quadruplex HMWs when

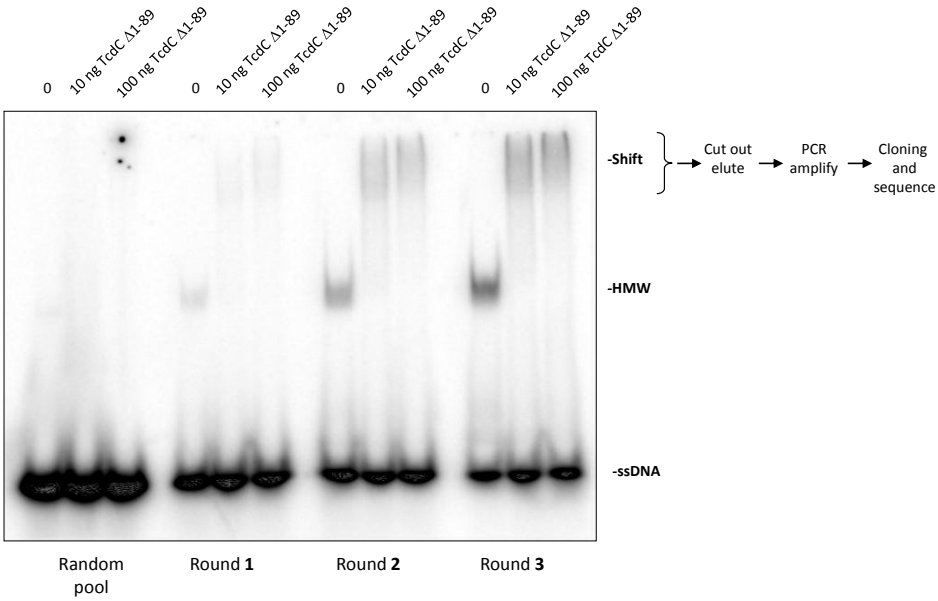

Figure 6. Mobility shift assay of TcdC Δ1-89 selected binding sites. ssDNA selected at each round (see 'Materials and Methods') were used as probes in gel mobility analysis. The selection rounds are indicated. Shifted probes in round 2 and 3 were cut out and eluted as indicated and cloned after the last round.

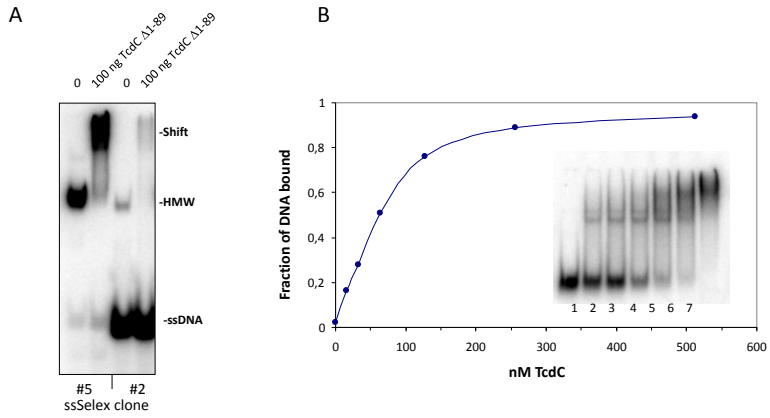

Figure 7.Binding characteristic of TcdC ⊿1-89. **A**. Binding to two selected clones (#2 and #5 corresponding to Table 3) was tested using 100 ng of TcdC⊿1-89 in a bandshift assay. ssDNA = single stranded; HMW= higher molecular weight product formed; shift indicated protein-DNA complex). **B**. Determination of the dissociation constant of TcdC⊿1-89 for #5 HMW binding. Lanes 1 to 7 of the inset gel shows increasing concentrations of TcdC⊿1-89 3 ng, 6 ng, 12 ng, 25ng, 50 ng and 100 ng. Band shift assays were performed as described under material and methods.

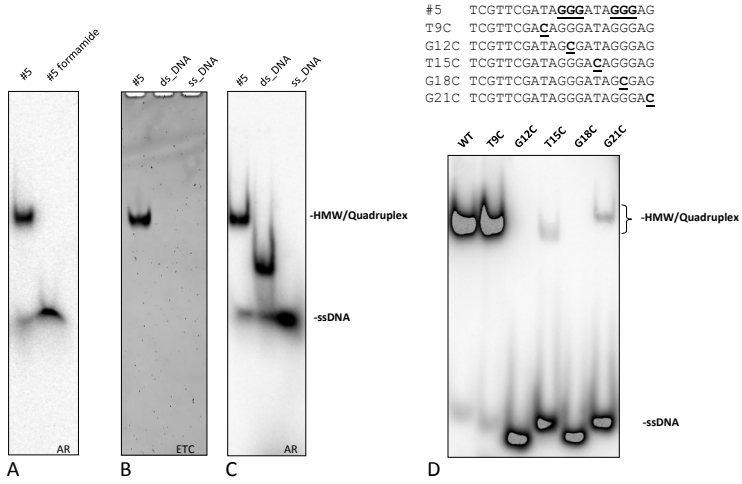

Figure 8: Characterization of a TcdC bound sequence element. **(A)** Higher molecular weight product (HMW) is formed by inter molecular interactions which are lost at 5 min 95°C in formamide. DNA was 5'end labeled with ^{32}Pγ-phosphate for visualization using storage phosphor screen autoradiography (AR). **(B)** Recognition of HMW product by ETC: a quadruplex specific stain (see 'Materials and Methods'). **(C)** Loading control of gel-samples loaded in panel B. Each DNA was 5'end labeled with ^{32}P-γ-phosphate see 'Materials and Methods'). **(D)** Point mutations affecting G-quadruplex formation. G-stretches and mutated positions are in bold underlined. Probes were ^{32}P-labelled and separated on a 10% polyacrylamide gel. Quadruplex is indicated.

the T9 preceding the G-stretches was replaced with a C (Figure 8D). However mutations at positions 12, 15, 18 and 21 of the #5 sequence resulted in significant alteration of migration on the gel, presumably due to a loss of QS. Especially when guanines involved in a G-tetrad formation, G12 and G15, were substituted with a C, the mutations completely abrogated the capacity of the sequence to fold into a G-quadruplex.

Table 3: Selected binding sites for the TcdCΔ1-89 protein in a single stranded SELEX.

#1	TCGGTGTGTT**GGG**TG**AGGG**AC
#2	TCGGCCTGGATACAT**AGGG**AC
#3	TCGGAATGACTGGCGT**GGG**AC
#4	TCGC**GGG**TGGCTGGA**AGGG**AC
#5	TCGTTCGAT**AGGG**AT**AGGG**AC
#6	TCGTTGTCTGGTC**AGGG**GGAC
#7	TCGAGCTATAGGT**GGG**TAGAC
#8	TCGGT**AGGGG**G**AGGG**A**GGG**AC
#9	TCGACAAAGCAT**GGG**TCCGAC
#10	TCGGTCTTTTG**GGG**TAAGGAC
#11	TCGTTTAGG**AGGG**TCTAGAC
#12	TCGAATATG**GGG**AAGTAGGAC
#13	TCGATTTG**GGG**ACTGCTGGAC
#14	TCGCGTC**AGG**AGGTGTTAGAC
#15	TCGCGG**AGGG**AAC**GGG**TGGAC
#16	TCGTAA**AGGG**TGATTCTGGAC
#17	TCGG**AGGG**CCAGGTCGTGAC
#18	TCG**AGGG**TTACCGTAG**GGG**AC
consensus	**aGGG**

Oligonucleotides sequences obtained are aligned. Stretches of 3G or longer and corresponding to the consensus are highlighted in bold. Underlined are the constant sequences flanking the randomized 15 nucleotides.

TcdC binds as a dimer

Above we have shown that the dimerization coiled-coil helix forms a proteolytically protected structure together with the OB-fold. We were interested to determine if dimerization is required for efficient recognition and binding of the G-quadruplex. Therefore electrophoretic mobility shift assay was carried out with TcdCΔ1-130, which behaves as a monomer. Figure 9 shows that no binding occurred with purified TcdCΔ1-130 indicating requirement of the TcdC dimerization domain for efficient binding. To exclude the possibility that the loss of binding by TcdCΔ1-130, is caused by direct of binding of the coiled-coil domain to the quadruplex we tested a TcdC protein, which does contain the dimerization helix (aa90-130) but misses the C-terminal part of the OB-fold containing a loop forming part of the putative

ssDNA-binding channel (aa208-232). Binding of this protein, TcdC Δ1-89_Δ208-232 (Figure 9), to the quadruplex structure was tested and showed no binding confirming that the coiled-coil domain is not directly involved in quadruplex binding.

Discussion

TcdC has been described to act as a factor responsible for inhibition of transcription of the toxin genes. Here we describe that the conserved carboxy terminal domain of TcdC is predicted to form a coiled five-stranded beta-sheet capped by an alpha helix (Figure 1A). This common fold has been described in different proteins which bind oligonucleotides or oligosaccharides and thus named OB-fold (oligonucleotide / oligosaccharide binding) (34). Using ssSELEX, a method to determine the binding site of TcdC, we found that the optimal binding site forms a G-quadruplex.

G-quadruplexes are nucleic acid sequences rich in guanine and capable of constituting a four-stranded structure. These four stranded structures are stabilized through hydrogen bonds between four guanine bases forming a square planar structure called a guanine tetrad (35). Typically three guanine tetrads can stack on top of each other to form a G-quadruplex. Quadruplexes can be formed in DNA as well as RNA and can be very diverse as GGG interactions can form intramolecularly as well as intermolecular. Intermolecular quadruplexes can be arranged from two strands (each containing two GGG stretches) or four strands (one GGG stretch each). The spacing (loops) between the GGG-stretches can vary between 1-7 nucleotides (36). *In silico* studies have shown that putative Quadruplex structures (pQS) are abundant in prokaryotic as well as eukaryotic gene promoters and G-rich telomeres found at the end of chromosomes (37-40). The presence of QS in human gene promoters has been shown to result in transcriptional repression (38). When QS are present in the 5' untranslated region of the mRNA it can interfere with ribosome binding and translation initiation (41;42).When we analyzed the whole *C.difficile* (strain 630) genome in Quadfinder, an online server for prediction of quadruplex-forming motifs in nucleotide sequences (36), we found 5 pQS. Unfortunately none of these were located in the PaLoc, where TcdC is speculated to act. Three pQS are present within open reading frames (CD1092A, CD1115 and CD1849) and two in the 3'UTR of genes (CD0938 and CD2929) in the strand that is complementary to mRNA (producing CCC stretches in the mRNA). It should be noted that prediction programs can only identify intramolecular QS (4 GGG stretches on the same strand) not the bimolecular or tetramolecular forms.

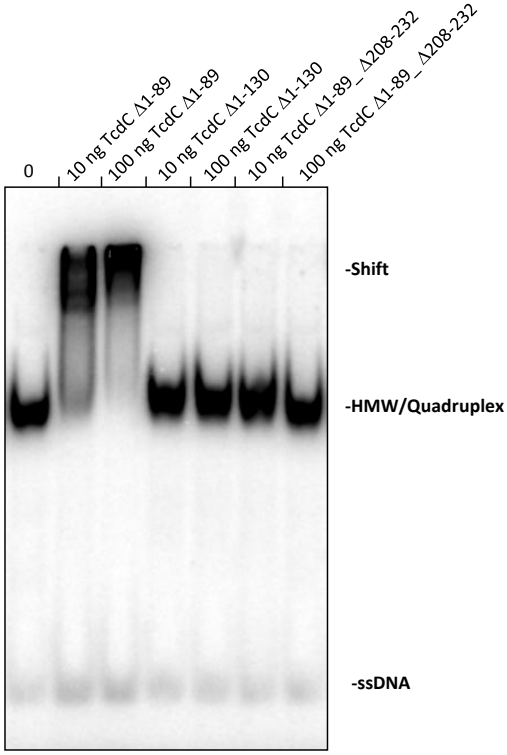

Figure 9: TcdC dimerization domain is required for DNA binding. TcdC dimer (Δ1-89) or monomer (Δ1-130) was incubated with oligo #5. Only the TcdC dimer was able to bind to the quadruplex DNA, as evidenced by the shifted DNA.

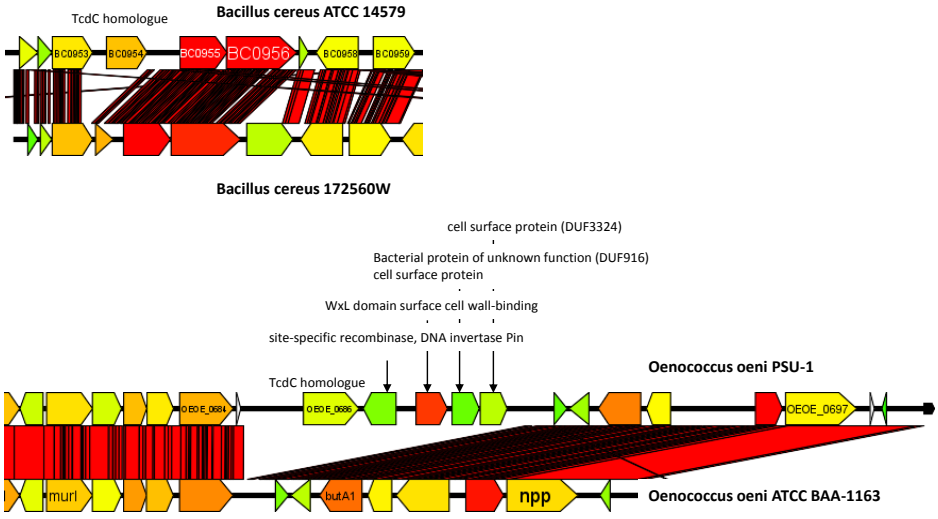

Figure 10: Pairwise comparison between TcdC homologues containing Firmicutes. Two examples are shown.Top; comparison between Bacillus cereus ATCC 14579 and Bacillus cereus 172560W showing a single insertion of the TcdC homologue BC0954. Bottom; comparison between Oenococcus oeni PSU-1 and Oenococcus oeni ATCC BAA-1163 showing an insertion of five genes including the TcdC homologue OEOE_068.

Telomers contain single stranded repeats of TAGGG found at the end of chromosomes protecting them from exonuclease degradation. Intramolecular QS of these TAGGG repeats play a role in telomere maintenance (43). However, in order to efficiently replicate the lagging strand of these telomeres these QS must be disrupted, thereby permitting processive telomere elongation. At least two proteins have been reported to bind and unfold these quadruplex structures, human POT1 and RPA, both characterized by the presence of an OB-fold (44-47). Despite the similarity in protein fold and DNA recognition site between TcdC (aGGG) and these OB-fold proteins (TAGGG), we could not find identical DNA contacting amino-acids when the TcdC structure was superimposed on the hPOT-1-DNA co-crystal (43). Although the circular genomes of Firmicutes do not contain single-stranded ends, a role for the OB-fold containing TcdC homologues in destabilizing alternative DNA secondary structures could be envisaged.

An *alternative* mechanism of TcdC action might be exemplified by the eukaryotic RNA polymerase II complex, which includes a subunit (rpb7) with an OB-fold. It was speculated that this OB-fold domain, which is located at the RNA exit path, binds RNA as it exits the enzyme thereby stabilizing the early transcribing complex (48). An opposite effect i.e. destabilizing the initiation complex by an OB-fold protein, such as TcdC, could be pictured. An unexpected G-quadruplex structure is described in the crystal structure of a bacterial -10 promoter element, 5'TGTACAATGGG 3' (-14 to -4), complexed with sigma factor Taqσ^A (49). In this structure the downstream $G_{-6}G_{-5}G_{-4}$ do not interact with the protein but twist away from the protein-DNA complex and form G-quadruplexes with other (symmetry-related) GGG motifs. The relevance in this complex was not clear and such a GGG-motif is absent next to the *tcdA* -10 promoter element.

TcdC is part of the pathogenicity locus, a well defined genetic element that is present at identical locations in the chromosome of pathogenic *C.difficile* strains. In non-toxinogenic strains, however, it is completely absent. These observations have led to the suggestion that the PaLoc may be associated with a (bacteriophage) transposable genetic element (50). Examining the genomic location of TcdC homologues of other Firmicutes showed that several of these family members are located on insertional-elements. For example the TcdC homologue (E-value 1e-23) of *Oenococussoeni* strain PSU-1 (Figure 10) is part of an insertion containing 4 additional genes encoding three putative cell-wall proteins and one site specific recombinase. In contrast, a TcdC homologue of *Bacillus cereus* strain 172560 (E-value 1e-41) is inserted without any additional genes.

It is interesting to mention that the TcdC variants present in *LactoBacillus* and *Leuconostoc* are also found in their homologous phages (i.e. *LactoBacillus* phages A2 and Lrm1 and *Leuconostoc* phage phiMH1).In these phages the TcdC homologues

are present in the lysis/lysogeny genetic switch operon, located between the CI repressor and Int, integrase, suggesting that TcdC is part of the regulatory decision circuit.

Our overall data suggests that *C.difficile* TcdC forms an OB-fold that binds quadruplex structures. However, the *in vivo* relevance remains unreported. Extensive investigations showed no binding to *tcdA* promoter elements. Clearly, quadruplex structures play a role in gene regulation and expression. Unfortunately, no multiple G-stretches are found within the PaLoc where TcdC is thought to exert its function. It may well be that the single stranded regions of the quadruplex mimics another structure bound by TcdC and the quadruplex is an approximation of the optimal structural binding determinant. It remains to be established in which way the capability of dimeric TcdC to bind G-quadruplexes demonstrated in this study relates to its role as a transcriptional repressor or another cellular function.

Reference List

1. Goorhuis,A., Bakker,D., Corver,J., Debast,S.B., Harmanus,C., Notermans,D.W., Bergwerff,A.A., Dekker,F.W. and Kuijper,E.J. (2008) Emergence of *Clostridium difficile* infection due to a new hypervirulent strain, polymerase chain reaction Ribotype 078. *Clin. Infect. Dis.*, 47, 1162-1170.
2. Kuijper,E.J., Barbut,F., Brazier,J.S., Kleinkauf,N., Eckmanns,T., Lambert,M.L., Drudy,D., Fitzpatrick,F., Wiuff,C., Brown,D.J. *et al.* (2008) Update of *Clostridium difficile* infection due to PCR Ribotype 027 in Europe, 2008. *Euro. Surveill*, 13.
3. McDonald,L.C., Killgore,G.E., Thompson,A., Owens,R.C., Jr., Kazakova,S.V., Sambol,S.P., Johnson,S. and Gerding,D.N. (2005) An epidemic, toxin gene-variant strain of *Clostridium difficile*. *N. Engl. J. Med.*, 353, 2433-2441.
4. Pepin,J., Valiquette,L. and Cossette,B. (2005) Mortality attributable to nosocomial *Clostridium difficile*-associated disease during an epidemic caused by a hypervirulent strain in Quebec. *CMAJ.*, 173, 1037-1042.
5. Rupnik,M., Wilcox,M.H. and Gerding,D.N. (2009) *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat. Rev. Microbiol.*, 7, 526-536.
6. Govind,R. and Dupuy,B. (2012) Secretion of *Clostridium difficile* toxins A and B requires the holin-like protein TcdE. *PLoS. Pathog.*, 8, e1002727.
7. Mani,N. and Dupuy,B. (2001) Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. *Proc. Natl. Acad. Sci. U. S. A*, 98, 5844-5849.
8. Raffestin,S., Dupuy,B., Marvaud,J.C. and Popoff,M.R. (2005) BotR/A and TetR are alternative RNA polymerase sigma factors controlling the expression of the neurotoxin and associated protein genes in *Clostridium* botulinum type A and *Clostridium* tetani. *Mol. Microbiol.*, 55, 235-249.
9. Mani,N., Lyras,D., Barroso,L., Howarth,P., Wilkins,T., Rood,J.I., Sonenshein,A.L. and Dupuy,B. (2002) Environmental response and autoregulation of *Clostridium difficile* TxeR, a sigma factor for toxin gene expression. *J. Bacteriol.*, 184, 5971-5978.
10. Antunes,A., Camiade,E., Monot,M., Courtois,E., Barbut,F., Sernova,N.V., Rodionov,D.A., Martin-Verstraete,I. and Dupuy,B. (2012) Global transcriptional control by glucose and carbon regulator CcpA in *Clostridium difficile*. *Nucleic Acids Res.*.
11. Dineen,S.S., Villapakkam,A.C., Nordman,J.T. and Sonenshein,A.L. (2007) Repression of *Clostridium difficile* toxin gene expression by CodY. *Mol. Microbiol.*, 66, 206-219.
12. Hughes,K.T. and Mathee,K. (1998) The anti-sigma factors. *Annu. Rev. Microbiol.*, 52, 231-286.
13. Govind,R., Vediyappan,G., Rolfe,R.D. and Fralick,J.A. (2006) Evidence that *Clostridium difficile* TcdC is a membrane-associated protein. *J. Bacteriol.*, 188, 3716-3720.
14. Bakker,D., Smits,W.K., Kuijper,E.J. and Corver,J. (2012) TcdC Does Not Significantly Repress Toxin Expression in *Clostridium difficile* 630DeltaErm. *PLoS. One.*, 7, e43247.
15. Merrigan,M., Venugopal,A., Mallozzi,M., Roxas,B., Viswanathan,V.K., Johnson,S., Gerding,D.N. and Vedantam,G. (2010) Human hypervirulent *Clostridium difficile* strains exhibit increased sporulation as well as robust toxin production. *J. Bacteriol.*, 192, 4904-4911.
16. Vohra,P. and Poxton,I.R. (2011) Comparison of toxin and spore production in clinically relevant strains of *Clostridium difficile*. *Microbiology*, 157, 1343-1353.
17. Matamouros,S., England,P. and Dupuy,B. (2007) *Clostridium difficile* toxin expression is inhibited by the novel regulator TcdC. *Mol. Microbiol.*, 64, 1274-1288.

18. Carter,G.P., Douce,G.R., Govind,R., Howarth,P.M., Mackin,K.E., Spencer,J., Buckley,A.M., Antunes,A., Kotsanas,D., Jenkin,G.A. *et al.* (2011) The anti-sigma factor TcdC modulates hypervirulence in an epidemic BI/NAP1/027 clinical isolate of *Clostridium difficile*. *PLoS. Pathog.*, 7, e1002317.
19. Cartman,S.T., Kelly,M.L., Heeg,D., Heap,J.T. and Minton,N.P. (2012) 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.
20. Yang,Q., Xiang,J., Yang,S., Li,Q., Zhou,Q., Guan,A., Zhang,X., Zhang,H., Tang,Y. and Xu,G. (2010) Verification of specific G-quadruplex structure by using a novel cyanine dye supramolecular assembly: II. The binding characterization with specific intramolecular G-quadruplex and the recognizing mechanism. *Nucleic Acids Res.*, 38, 1022-1033.
21. Wolf,E., Kim,P.S. and Berger,B. (1997) MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci.*, 6, 1179-1189.
22. Roy,A., Kucukural,A. and Zhang,Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.*, 5, 725-738.
23. Zhang,Y. (2007) Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins*, 69 Suppl 8, 108-117.
24. Hildebrand,A., Remmert,M., Biegert,A. and Soding,J. (2009) Fast and accurate automatic structure prediction with HHpred. *Proteins*, 77 Suppl 9, 128-132.
25. Arcus,V. (2002) OB-fold domains: a snapshot of the evolution of sequence, structure and function. *Curr. Opin. Struct. Biol.*, 12, 794-801.
26. Theobald,D.L., Mitton-Fry,R.M. and Wuttke,D.S. (2003) Nucleic acid recognition by OB-fold proteins. *Annu. Rev. Biophys. Biomol. Struct.*, 32, 115-133.
27. Soding,J., Biegert,A. and Lupas,A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.*, 33, W244-W248.
28. Chen,J., Darst,S.A. and Thirumalai,D. (2010) Promoter melting triggered by bacterial RNA polymerase occurs in three steps. *Proc. Natl. Acad. Sci. U. S. A*, 107, 12523-12528.
29. Wigneshweraraj,S.R., Burrows,P.C., Severinov,K. and Buck,M. (2005) Stable DNA opening within open promoter complexes is mediated by the RNA polymerase beta'-jaw domain. *J. Biol. Chem.*, 280, 36176-36184.
30. McGlynn,P., Mahdi,A.A. and Lloyd,R.G. (2000) Characterisation of the catalytically active form of RecG helicase. *Nucleic Acids Res.*, 28, 2324-2332.
31. Rafferty,J.B., Ingleston,S.M., Hargreaves,D., Artymiuk,P.J., Sharples,G.J., Lloyd,R.G. and Rice,D.W. (1998) Structural similarities between Escherichia coli RuvA protein and other DNA-binding proteins and a mutational analysis of its binding to the holliday junction. *J. Mol. Biol.*, 278, 105-116.
32. Svobodova,M., Pinto,A., Nadal,P. and O' Sullivan,C.K. (2012) Comparison of different methods for generation of single-stranded DNA for SELEX processes. *Anal. Bioanal. Chem.*, 404, 835-842.
33. Xu,D., Guo,R., Sobeck,A., Bachrati,C.Z., Yang,J., Enomoto,T., Brown,G.W., Hoatlin,M.E., Hickson,I.D. and Wang,W. (2008) RMI, a new OB-fold complex essential for Bloom syndrome protein to maintain genome stability. *Genes Dev.*, 22, 2843-2855.
34. Murzin,A.G. (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.*, 12, 861-867.

35. Bochman,M.L., Paeschke,K. and Zakian,V.A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.*, 13, 770-780.
36. Scaria,V., Hariharan,M., Arora,A. and Maiti,S. (2006) Quadfinder: server for identification and analysis of quadruplex-forming motifs in nucleotide sequences. *Nucleic Acids Res.*, 34, W683-W685.
37. Lipps,H.J. and Rhodes,D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.*, 19, 414-422.
38. Qin,Y. and Hurley,L.H. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie*, 90, 1149-1171.
39. Rawal,P., Kummarasetti,V.B., Ravindran,J., Kumar,N., Halder,K., Sharma,R., Mukerji,M., Das,S.K. and Chowdhury,S. (2006) Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation. *Genome Res.*, 16, 644-655.
40. Yadav,V.K., Abraham,J.K., Mani,P., Kulshrestha,R. and Chowdhury,S. (2008) QuadBase: genome-wide database of G4 DNA--occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. *Nucleic Acids Res.*, 36, D381-D385.
41. Bugaut,A. and Balasubramanian,S. (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.*, 40, 4727-4741.
42. Wieland,M. and Hartig,J.S. (2007) RNA quadruplex-based modulation of gene expression. *Chem. Biol.*, 14, 757-763.
43. Lei,M., Podell,E.R., Baumann,P. and Cech,T.R. (2003) DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA. *Nature*, 426, 198-203.
44. Paeschke,K., Simonsson,T., Postberg,J., Rhodes,D. and Lipps,H.J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.*, 12, 847-854.
45. Salas,T.R., Petruseva,I., Lavrik,O., Bourdoncle,A., Mergny,J.L., Favre,A. and Saintome,C. (2006) Human replication protein A unfolds telomeric G-quadruplexes. *Nucleic Acids Res.*, 34, 4857-4865.
46. Wu,L., Multani,A.S., He,H., Cosme-Blanco,W., Deng,Y., Deng,J.M., Bachilo,O., Pathak,S., Tahara,H., Bailey,S.M. *et al.* (2006) Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres. *Cell*, 126, 49-62.
47. Zaug,A.J., Podell,E.R. and Cech,T.R. (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc. Natl. Acad. Sci. U. S. A*, 102, 10864-10869.
48. Spahr,H., Calero,G., Bushnell,D.A. and Kornberg,R.D. (2009) Schizosacharomyces pombe RNA polymerase II at 3.6-A resolution. *Proc. Natl. Acad. Sci. U. S. A*, 106, 9185-9190.
49. Feklistov,A. and Darst,S.A. (2011) Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell*, 147, 1257-1269.
50. Braun,V., Hundsberger,T., Leukel,P., Sauerborn,M. and von Eichel-Streiber,C. (1996) Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*. *Gene*, 181, 29-38.