



## A Catalytic DNA Activated by a Specific Strain of Bacterial Pathogen

Zhifa Shen<sup>+</sup>, Zaisheng Wu<sup>+</sup>, Dingran Chang<sup>+</sup>, Wenqing Zhang, Kha Tram, Christine Lee, Peter Kim, Bruno J. Salena, and Yingfu Li\*

**Abstract:** Pathogenic strains of bacteria are known to cause various infectious diseases and there is a growing demand for molecular probes that can selectively recognize them. Here we report a special DNAzyme (catalytic DNA), RFD-CD1, that shows exquisite specificity for a pathogenic strain of *Clostridium difficile* (*C. difficile*). RFD-CD1 was derived by an in vitro selection approach where a random-sequence DNA library was allowed to react with an unpurified molecular mixture derived from this strain of *C. difficile*, coupled with a subtractive selection strategy to eliminate cross-reactivities to unintended *C. difficile* strains and other bacteria species. RFD-CD1 is activated by a truncated version of TcdC, a transcription factor, that is unique to the targeted strain of *C. difficile*. Our study demonstrates for the first time that in vitro selection offers an effective approach for deriving functional nucleic acid probes that are capable of achieving strain-specific recognition of bacterial pathogens.

**B**acterial infections pose serious threats to public health and are responsible for many annual costly outbreaks.<sup>[1]</sup> The issue has been further compounded by the emergence of hyper-virulent and/or antibiotic-resistant strains, which has now become a serious global problem.<sup>[2]</sup> Early detection of specific pathogens has long been recognized as a vital strategy in the control of infectious diseases because it can lead to timely care of patients and prevent potential outbreaks. However, detection of specific bacteria represents a significant challenge because of the presence of many different species of bacteria in biological samples. Furthermore, for any given species of bacterium, only virulent strains are infectious while other strains of the same species may be harmless or even beneficial to human health. Therefore, there is a great need to develop molecular probes that are highly specific for pathogenic strains of bacteria.

DNAzymes refer to single-stranded DNA molecules with catalytic capabilities.<sup>[3]</sup> They have been widely explored as molecular tools for applications ranging from biosensing to gene regulation.<sup>[4]</sup> DNAzymes can be generated de novo by in vitro selection, a simple technique that allows for the isolation of rare functional DNA sequences from a random DNA pool.<sup>[5]</sup> We have previously developed an RNA-cleaving fluorogenic DNAzyme (RFD) as an indicator for *E. coli*.<sup>[6]</sup> RFD-EC1 has since been used in a colorimetric assay for the field detection of *E. coli* and in a microfluidic-based counting assay for the rapid detection of *E. coli* in blood.<sup>[7,8]</sup> Although RFD-EC1 is highly specific for *E. coli* with minimal cross-activity to other bacterial species,<sup>[6]</sup> it non-discriminatively recognizes all *E. coli* strains.<sup>[8]</sup> The goal of this study was to investigate whether it was possible to derive strain-specific RFDs by in vitro selection.

*C. difficile*, a gram-positive bacterium that has been identified as the major cause of the diarrheic disease known as *Clostridium difficile* infection (CDI), was chosen for the current study. The incidence and mortality of CDI have increased dramatically over the past 15 years and CDI has become one of the most common healthcare-associated infections in the Western hemisphere.<sup>[9]</sup> These have been linked to the emergence of a hypervirulent clinical strain known as BI/027.<sup>[10]</sup> In addition to being more virulent, this strain is also more resistant to antibiotics that are used to treat CDI.<sup>[11]</sup> Therefore, developing a specific molecular probe for BI/027 has important clinical implications.

We used a locally isolated clinical strain of *C. difficile* for our investigation, which was confirmed to be a BI/027 strain by the typing experiment shown in Figure S1 in the Supporting Information. We refer this strain as BI/027-H (H: Hamilton, Ontario). We employed two strategies to derive the desired DNAzymes. First, we used a complex mixture (rather than a defined target) derived from BI/027-H for the

[\*] Dr. Z. Shen,<sup>[+]</sup> Dr. Z. Wu,<sup>[+]</sup> D. Chang,<sup>[+]</sup> W. Zhang, Dr. K. Tram, Prof. Dr. Y. Li  
Department of Biochemistry and Biomedical Sciences  
Michael G. DeGroote Institute for Infectious Disease Research  
McMaster University, 1280 Main Street West, Hamilton  
Ontario, L8S 4K1 (Canada)  
E-mail: liying@mcmaster.ca  
Prof. Dr. C. Lee  
Department of Pathology and Molecular Medicine  
St Joseph's Healthcare, McMaster University  
50 Charlton Ave E, 424 Luke Wing, Hamilton  
Ontario L8N 4A6 (Canada)  
Prof. Dr. B. J. Salena  
Department of Medicine, McMaster University  
1280 Main St. W., Hamilton  
Ontario L8S 4K1 (Canada)

Prof. Dr. P. Kim  
Department of Mathematics and Statistics  
University of Guelph, Guelph, Ontario N1G 2W1 (Canada)  
Dr. Z. Shen<sup>[+]</sup>  
Current address:  
School of Laboratory Medicine and Life Sciences  
Wenzhou Medical University, Wenzhou, Zhejiang 325035 (China)

[+] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201510125>.

© 2015 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

selection—this strategy bypassed the challenging process of identifying a biomarker as the starting point. Second, in each round of selection, we incorporated a counter-selection step where the DNA pool was selected against similar mixtures from control bacteria—this strategy was to ensure the desired recognition specificity. Three bacteria were chosen for this purpose, *E. coli*, *B. subtilis* (representing two different species of bacteria), and CD630 (representing a non-BI/027 *C. difficile* strain).

RFD isolation was achieved using a DNA library (Figure 1, panel B) containing 40 random nucleotides and a selection scheme shown in Figure S2. Prior to the selection, a crude extracellular mixture (CEM) was prepared for BI/027-H. This was done by culturing BI/027-H to OD<sub>600</sub> (optical density at 600 nm) of approximately 1, followed by cell

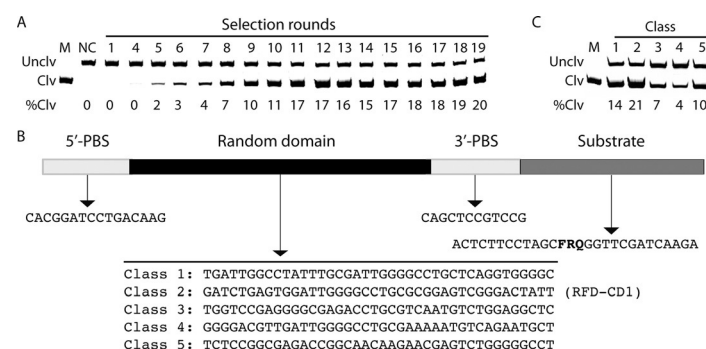
We examined the recognition specificity of RFD-CD1. We first evaluated the reactivity of RFD-CD1 towards the CEM of 13 bacterial species and found none of them was able to activate RFD-CD1 (Figure 2 A). We then tested the reactivity of RFD-CD1 to the CEM of 12 *C. difficile* strains. Only the CEM from BI/027 was able to activate RFD-CD1 (lane 1, Figure 2 B). The results indicate that RFD-CD1 is both species-specific and strain-selective.

Next we turned our attention to the identification of the target that activates RFD-CD1. We first treated CEM-CD with three proteases, trypsin (TP) and proteinase K (PK) and subtilisin (SL), to determine if the activator is a protein. We found that CEM-CD treated with each protease was unable to activate RFD-CD1 (Figure 3 A), signifying that the target is indeed a protein.

We evaluated the molecular weight of the target using molecular sizing. CEM-CD was passed through centrifugal filters with molecular weight (MW) cut-offs from 3–100 kilodaltons (kDa). The 3 kDa and 10 kDa filtrates did not induce the cleavage of RFD-CD1 while the 30 kDa filtrate caused a small level of cleavage. In contrast, the 50 kDa and 100 kDa filtrates prompted strong cleavage (Figure 3 B). This experiment suggests that the protein target has a MW of about 30 000 Daltons.

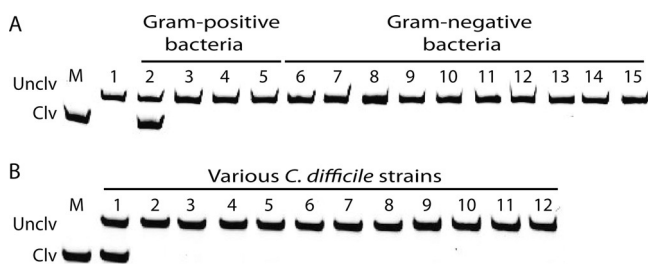
*C. difficile* genome encodes nearly 4000 proteins.<sup>[12]</sup> Given the estimated MW of the target, we tightened our search to 986 proteins with an MW between 25 000 and 35 000 Daltons (Figure 3 C). Based on the fact that the RFD-CD1 is a DNA molecule, we hypothesized that the target might have an intrinsic ability to interact with DNA. Therefore, we decided to narrow down our search to gene expression regulators. After applying this filtration strategy using Procom software, the potential candidates were further reduced to 78 (Figure 3 C).

The strain specificity data presented in Figure 2 B indicates that RFD-CD1 is only active with BI/027-H but not with other strains including CD630. Previous studies have shown that there are numerous genetic differences between CD-BI/027 and CD630.<sup>[13]</sup> Taking these into account, we decided to compare the genomic

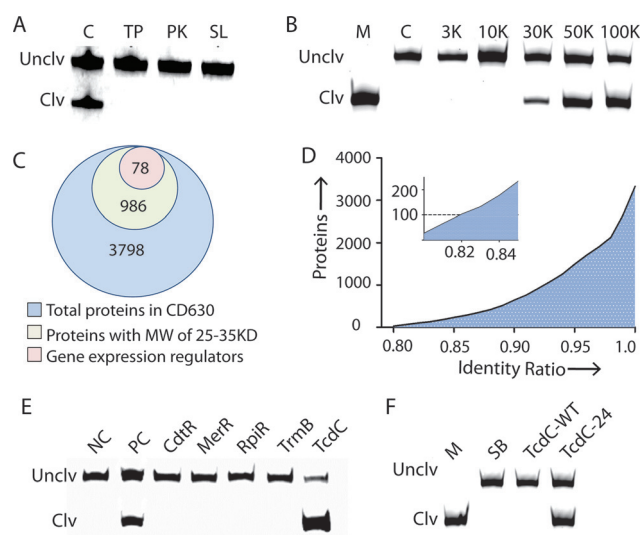


**Figure 1.** A) The activity of DNA pools. Each pool was incubated with CEM-CD for 30 minutes, followed by dPAGE analysis. Uncv: uncleaved DNA pool; Clv: cleaved DNA pool; %Clv =  $[F_{\text{Clv}}/6]/[(F_{\text{Clv}}/6) + F_{\text{Uncv}}]$ .  $F_{\text{Clv}}$  and  $F_{\text{Uncv}}$ : fluorescence intensity of cleaved and uncleaved fractions of the DNA pool. Note that the cleavage leads to six-fold fluorescence enhancement, which was taken into consideration for cleavage percentage calculation. M: marker for cleavage. NC: Negative control (pool 19 in the selection buffer). B) Five RFD classes. Each sequence is organized in the 5'-3' direction as 5'-PBS (primer binding site), random domain, 3'-PBS, and substrate. F: fluorescein-dT; R: adenine ribonucleotide; Q: DabcyI-dT. C) Responses of top RFDs towards CEM-CD. Reaction time: 30 minutes.

removal using centrifugation. This CEM, named CEM-CD, served as the positive selection target. A CEM was also prepared similarly for each control bacterium. These CEMs were combined and incubated with the DNA library for five hours. The uncleaved DNA molecules in this step were purified by 10% denaturing polyacrylamide gel electrophoresis (dPAGE), and then incubated with CEM-CD for 30 minutes. The cleaved DNA products in this step were purified by dPAGE, amplified by polymerase chain reaction (PCR), and used for the next round of selection (experimental details are described in the Supporting Information). In total, 19 iterations were conducted. Figure 1 A depicts the reactivity of representative DNA pools when incubated with CEM-CD. By round 19, approximately 20% cleavage was observed. The 19th DNA pool was cloned and sequenced. Five DNAzyme classes were discovered (Figure 1 B) and the DNAzyme with highest cleavage activity (class 2) was named RFD-CD1 (Figure 1 C) and chosen for further investigation.



**Figure 2.** A) Responses of RFD-CD1 to CEMs prepared from various bacteria. Lane 1: reaction buffer only; lanes 2–5: *C. difficile*, *B. subtilis*, *Leuconostoc mesenteroides*, and *Pediococcus acidilactici*. Lanes 6–15: *Pseudomonas peli*, *Brevundimonas diminuta*, *Hafnia alvei*, *Yersinia ruckeri*, *Ochrobactrum grignone*, *Achromobacter xylosoxidans*, *Moraxella osloensis*, *Acinetobacter lwoffii*, *Serratia fonticola*, and *E. coli*. B) Responses of RFD-CD1 to CEMs prepared from different *C. difficile* strains. Lane 1: BI/027-H; Lanes 2–12: BAA-1801, BAA-1804, BAA-1875, 43594, 43598, BAA-1382 (also known as CD630), BAA-1871, BAA-1872, 43255, BAA-1870, and BAA-1814, respectively. Reaction time: 30 minutes.



**Figure 3.** A) Responses of RFD-CD1 to protease-treated CEM-CD. TP: trypsin; PK: proteinase K; SL: subtilisin. B) Estimation of the molecular weight of the protein target. CEM-CD was passed through centrifugal filters with specified molecular weight cut-offs, and the filtrates were then tested for reactivity with RFD-CD1. C) The Venn diagram showing the number of proteins in CD630 in each listed category. D) Cumulative distribution of proteins in CD630 and CD196 having lowest to highest identity ratio. Insert shows the first 100 proteins with the lowest identity ratio. E) RFD-CD1 treated with the CEM prepared from *E. coli* cells transformed with a plasmid expressing one of the five candidate targets: CdtR, MerR, RpiR, TrmB, and TcdC. NC: the negative control made of the CEM prepared from *E. coli* containing an empty plasmid. PC: positive control made of the CEM prepared from BI/027-H. F) RFD-CD1 treated with purified TcdC-WT and TcdC-24. SB: Selection buffer only. Reaction time in panels A, B, E and F: 30 minutes.

difference between CD630 and CD196, a BI/027 strain whose genomic sequence has been reported.<sup>[13]</sup>

We calculated the sequence identity of each protein encoded by a given gene in these two strains by Blastx<sup>[14]</sup> and the results are presented in Figure 3D. Upon examination of the list of the top 100 proteins with lowest sequence identity with the list of the 78 candidates described earlier, we found that the following five proteins appeared on both: MerR, TcdC, RpiR, TrmB, and CdtR. Each candidate gene was then amplified from the genomic DNA of BI/027-H and expressed in *E. coli*. The CEM from each cell line was then prepared and used to induce the cleavage of RFD-CD1. Of the five cell lines, only the one harboring the *tcdC* gene produced a positive signal (Figure 3E), indicating that TcdC, a transcription factor, is the target for RFD-EC1.

To further verify that TcdC is the target for RFD-CD1 and uncover the reason behind the strain specificity, we sequenced the *tcdC* genes in BI/027-H and CD630 (Table 1). The *tcdC* gene from CD630 contains an open reading frame (ORF) of 696 nucleotides, translating into a TcdC protein that has a size of 232 amino acids, which is denoted TcdC-WT.<sup>[15]</sup> The nucleotide and protein sequences match those in the NCBI databank.<sup>[16]</sup> The *tcdC* gene from BI/027-H, however, has an internal truncation of 18 nucleotides located between 330–347. The same gene has been reported previously by others,

**Table 1:** Typing of *tcdC* genes of the *C. difficile* strains used in Figure 2 B for strain specificity test.

| Strain names     | TcdC version | Size of TcdC |
|------------------|--------------|--------------|
| BI/027-H         | TcdC-24      | 226          |
| BAA-1382 (CD630) | TcdC-WT      | 232          |
| 43 255           | TcdC-WT      | 232          |
| 43 598           | TcdC-WT      | 232          |
| 43 594           | TcdC-WT      | 232          |
| BAA-1804         | TcdC-WT      | 232          |
| BAA-1814         | TcdC-WT      | 232          |
| BAA-1871         | TcdC-WT      | 232          |
| BAA-1872         | TcdC-WT      | 232          |
| BAA-1870         | TcdC-1       | 65           |
| BAA-1875         | TcdC-5       | 61           |
| BAA-1801         | no TcdC      | 0            |

which was named *tcdC-24*<sup>[15]</sup> (or *tcdC-B*<sup>[17]</sup>). Thus, the TcdC protein from BI/027-H, TcdC-24, is 6-amino acid shorter than TcdC-WT.

We cloned and purified TcdC-WT and TcdC-24, which are predicted to have a molecular weight of 28200 and 27300 Daltons, respectively. These predictions are consistent with SDS-PAGE analysis showing that TcdC-24 has an increased gel mobility (Figure S3). Purified TcdC-24 and TcdC-WT were then used to induce the cleavage of RFD-CD1. As expected, TcdC-24 induced the cleavage of RFD-CD1 while TcdC-WT did not (Figure 3F). These experiments confirm that TcdC is indeed the target for RFD-CD1 and explain the specificity of RFD-CD1 for BI/027-H over CD630.

We next sequenced the *tcdC* genes from other strains used for the strain specificity test in Figure 2 B and the results are summarized in Table 1. BAA-1801 is a non-pathogenic strain and does not have a *tcdC* gene.<sup>[18]</sup> BAA-1870 and BAA-1875 produce significantly truncated TcdC proteins containing only 65 and 61 amino acids, respectively, due to sequence mutations.<sup>[15]</sup> All other *C. difficile* strains produce wild-type TcdC. Taken together, the experiments above not only result in the discovery of TcdC as the target for RFD-CD1, but also explain the origin of the strain specificity. That is: RFD-CD1 has a high specificity for TcdC-24 and recognizes neither full-length TcdC nor the truncated TcdC variants.

In summary, using in vitro selection, we have obtained an RNA-cleaving fluorogenic DNAzyme (RFD) that can recognize an infectious strain of *C. difficile*. This DNAzyme not only exhibits no cross-reactivity to other bacterial species, but also is highly strain-selective for *C. difficile*. To our knowledge, our work reports the first example of a functional nucleic acid that can recognize a specific strain of bacterium. Moreover, our approach also represents an unconventional strategy to molecular probe engineering as it does not begin with the identification of a specific biomarker for the probe development, which can be a lengthy and possibly unproductive process. Instead, our approach relies on a test-tube selection process to come up with a solution. This is achieved through tandem selection steps where a complex biological mixture from the targeted bacterial strain is used as the positive-selection target and similar mixtures from control bacteria are used as the subtractive selection target. Even though this approach does present the challenge of identifying the target

upon the completion of in vitro selection, we can exploit simple bioinformatic tools and experimental techniques to progressively narrow down target candidates and ultimately identify the target. Taken together, our work here demonstrates that in vitro selection can be an effective solution to engineering functional DNA probes that are able to recognize specific strains of bacterial pathogens. This sets up the stage to exploit synthetic DNAzyme probes for infectious disease diagnosis.

### Acknowledgements

Funding is provided by Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institute of Health Research (CIHR) through a CHRP grant.

**Keywords:** aptamers · bacterial detection · biosensors · DNAzymes · in vitro selection

**How to cite:** *Angew. Chem. Int. Ed.* **2016**, *55*, 2431–2434  
*Angew. Chem.* **2016**, *128*, 2477–2480

- 
- [1] a) R. M. Kleven, J. R. Edwards, C. L. Richards, Jr., T. C. Horan, R. P. Gaynes, D. A. Pollock, D. M. Cardo, *Public Health Rep.* **2007**, *122*, 160–166; b) P. S. Mead, L. Slutsker, V. Dietz, L. F. McCaig, J. S. Bresee, C. Shapiro, P. M. Griffin, R. V. Tauxe, *Emerging Infect. Dis.* **1999**, *5*, 607–625.
- [2] Antibiotic resistance threats in the United States, **2013**. Centers for Disease Control and Prevention. It can be accessed at: <http://www.cdc.gov/drugresistance/threat-report-2013/>.
- [3] a) R. R. Breaker, *Nat. Biotechnol.* **1997**, *15*, 427–431; b) K. Schlosser, Y. Li, *Chem. Biol.* **2009**, *16*, 311–322.
- [4] a) J. Liu, Z. Cao, Y. Lu, *Chem. Rev.* **2009**, *109*, 1948–1998; b) S. K. Silverman, *Angew. Chem. Int. Ed.* **2010**, *49*, 7180–7201; *Angew. Chem.* **2010**, *122*, 7336–7359.
- [5] a) A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 818–822; b) C. Tuerk, L. Gold, *Science* **1990**, *249*, 505–510.
- [6] M. M. Ali, S. D. Aguirre, H. Lazim, Y. Li, *Angew. Chem. Int. Ed.* **2011**, *50*, 3751–3754; *Angew. Chem.* **2011**, *123*, 3835–3838.
- [7] K. Tram, P. Kanda, B. J. Salena, S. Huan, Y. Li, *Angew. Chem. Int. Ed.* **2014**, *53*, 12799–12802; *Angew. Chem.* **2014**, *126*, 13013–13016.
- [8] D. K. Kang, M. M. Ali, K. Zhang, S. S. Huang, E. Peterson, M. A. Digman, E. Gratton, W. Zhao, *Nat. Commun.* **2014**, *5*, 5427.
- [9] a) J. G. Bartlett, N. Moon, T. W. Chang, N. Taylor, A. B. Onderdonk, *Gastroenterology* **1978**, *75*, 778–782; b) H. E. Larson, A. B. Price, P. Honour, S. P. Borriello, *Lancet* **1978**, *311*, 1063–1066; c) R. Voelker, *JAMA J. Am. Med. Assoc.* **2010**, *303*, 2017–2019.
- [10] a) A. Goorhuis, T. Van der Kooi, N. Vaessen, F. W. Dekker, R. Van den Berg, C. Harmanus, S. van den Hof, D. W. Notermans, E. J. Kuijper, *Clin. Infect. Dis.* **2007**, *45*, 695–703; b) V. G. Loo, et al., *N. Engl. J. Med.* **2005**, *353*, 2242–2249; c) M. D. Redelings, F. Sorvillo, L. Mascola, *Emerging Infect. Dis.* **2007**, *13*, 1417–1419; d) G. Killgore, et al., *J. Clin. Microbiol.* **2008**, *46*, 431–437.
- [11] S. T. Cartman, J. T. Heap, S. A. Kuehne, A. Cockayne, N. P. Minton, *Int. J. Med. Microbiol.* **2010**, *300*, 387–395.
- [12] M. Sebahia, et al., *Nat. Genet.* **2006**, *38*, 779–786.
- [13] R. A. Stabler, et al., *Genome Biol.* **2009**, *10*, R102.
- [14] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**, *215*, 403–410.
- [15] K. E. Dingle, D. Griffiths, X. Didelot, J. Evans, A. Vaughan, *PLoS One* **2011**, *6*, e19993.
- [16] A. Samie, C. L. Obi, J. Franasiak, L. Archbald-Pannone, P. O. Bessong, C. Alcantara-Warren, R. L. Guerrant, *Am. J. Trop. Med. Hyg.* **2008**, *78*, 577–585.
- [17] P. Spigaglia, P. Mastrantonio, *J. Clin. Microbiol.* **2002**, *40*, 3470–3475.
- [18] L. Lemee, A. Dhalluin, S. Testelin, M. A. Mattrat, K. Maillard, J. F. Lemeland, J. L. Pons, *J. Clin. Microbiol.* **2004**, *42*, 5710–5714.

Received: October 30, 2015

Published online: December 16, 2015