Crystal structure and DNA binding activity of a PadR family transcription regulator from hypervirulent *Clostridium difficile* R20291

Catherine E. Isom\(^1,2\), Smita K. Menon\(^2\), Leonard M. Thomas\(^2\), Ann H. West\(^2\), George B. Richter-Addo\(^2\) and Elizabeth A. Karr\(^1,\ast\)

**Abstract**

**Background:** *Clostridium difficile* is a spore-forming obligate anaerobe that can remain viable for extended periods, even in the presence of antibiotics, which contributes to the persistence of this bacterium as a human pathogen during host-to-host transmission and in hospital environments. We examined the structure and function of a gene product with the locus tag CDR20291\(_{0991}\) (*cd*PadR\(_1\)) as part of our broader goal aimed at elucidating transcription regulatory mechanisms involved in virulence and antibiotic resistance of the recently emergent hypervirulent *C. difficile* strain R20291. *cd*PadR\(_1\) is genomically positioned near genes that are involved in stress response and virulence. In addition, it was previously reported that *cd*PadR\(_1\) and a homologue from the historical *C. difficile* strain 630 (CD630\(_{1154}\)) were differentially expressed when exposed to stressors, including antibiotics.

**Results:** The crystal structure of *cd*PadR\(_1\) was determined to 1.9 Å resolution, which revealed that it belongs to the PadR-s2 subfamily of PadR transcriptional regulators. *cd*PadR\(_1\) binds its own promoter and other promoter regions from within the *C. difficile* R20291 genome. DNA binding experiments demonstrated that *cd*PadR\(_1\) binds a region comprised of inverted repeats and an AT-rich core with the predicted specific binding motif, GTACTAT\(_{\text{N}2}\)ATTATA\(_{\text{N}}\)AGTA, within its own promoter that is also present in 200 other regions in the *C. difficile* R20291 genome. Mutation of the highly conserved W in \(\alpha_4\) of the effector binding/oligomerization domain, which is predicted to be involved in multi-drug recognition and dimerization in other PadR-s2 proteins, resulted in alterations of *cd*PadR\(_1\) binding to the predicted binding motif, potentially due to loss of higher order oligomerization.

**Conclusions:** Our results indicate that *cd*PadR\(_1\) binds a region within its own promoter consisting of the binding motif GTACTAT\(_{\text{N}2}\)ATTATA\(_{\text{N}}\)AGTA and seems to associate non-specifically with longer DNA fragments *in vitro*, which may facilitate promoter and motif searching. This suggests that *cd*PadR\(_1\) acts as a transcriptional auto-regulator, binding specific sites within its own promoter, and is part of a broad gene regulatory network involved, in part, with environmental stress response, antibiotic resistance and virulence.

**Keywords:** *Clostridium difficile*, Transcription regulation, Bacteria, Stress response, Structural biology, PadR

---

* Correspondence: lizkarr@ou.edu
1 Price Family Foundation Institute of Structural Biology and Department of Microbiology and Plant Biology, University of Oklahoma, 770 Van Vleet Oval, Norman, OK 73019, USA
Full list of author information is available at the end of the article

© 2016 The Author(s). Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Background
Epidemiological trends indicate clinical acquisition of *Clostridium difficile* as the primary route of human infection by this bacterium [1]. The risk of *C. difficile* becoming a community-acquired infection is likely to increase without the development of better identification and more effective treatment [2]. The genome of *C. difficile* has been described as “highly dynamic” based on the prevalence of horizontal gene transfer [3]. The impact of a genome that readily changes in response to environmental stress could be a major indicator of *C. difficile* pathogenicity [3]. *C. difficile* produces spores that allow it to be viable for extended periods, even in the presence of antibiotics, which could explain the persistence of this human pathogen during host-to-host transmission and in the hospital environment [4]. Transcription factors orchestrate the regulation of survival, proliferation, virulence, and antibiotic resistance mechanisms of human pathogens. As part of our larger goal aimed at elucidating structure and function of transcription regulatory mechanisms involved in virulence and antibiotic resistance of human pathogens, we focused on protein targets from a hypervirulent strain of *C. difficile* (R20291). Herein, we present our results on a member of the PadR family of transcription regulators (product of CDR20291_0991) that we have named cdPadR1.

The first described PadR proteins are transcriptional repressors for genes encoding phenolic acid decarboxylase (*padC*) that de-repress *padC* when phenolic acids are present in toxic amounts [5]. The PadR transcription regulator from *Bacillus subtilis* is a prototypical PadR-family member protein that binds the *padC* promoter in the absence of phenolic acid *in vitro*; binding is lost when exposed to phenolic acids [6, 7]. Unlike the prototypical PadR, the PadR-family transcription regulators AphA [8], LmrR [9], and bcPadR [10] from *Vibrio cholerae, Lactococcus lactis,* and *Bacillus cereus,* respectively, are involved in the regulation of virulence and antibiotic efflux mechanisms. The prototypical PadR and the PadR-like transcription regulator AphA are within a subfamily of PadR proteins (PadR-s1) which contain multiple α-helices in the C-terminal domain [10]. Another, less studied subfamily of PadR family proteins (PadR-s2) contains a single α-helix in the C-terminal effector binding/oligomerization domain [10]. The PadR-s2 proteins, which include the bcPadRs [10] and LmrR [11], have been structurally characterized and are involved in multiple drug recognition. The BC4206 gene product, bcPadR1, was upregulated 8.7-fold in the presence of enterocin treatment in *B. cereus* ATCC14572 when compared to an untreated control [12]. This PadR-like protein binds its own promoter and that of the gene BC4207, which encodes a membrane protein predicted to be involved in enterocin AS-48 resistance [12]. Binding of bcPadR1 to the predicted promoter region was not affected by the addition of AS-48 in vitro [10]. The PadR-like family protein of *L. lactis,* LmrR, binds the promoter region of an ABC-type multidrug transporter, LmrCD, and interacts with the compound Hoechst 33342 and the antibiotic daunomycin [9]. The crystal structure of apo-LmrR revealed a hydrophobic pore between αi of the dimer mates [11]. Additional structures of LmrR bound to Hoechst 33342 and daunomycin, separately, demonstrated that this pore is integral to inhibitor interaction [11]. The conformational change instigated at α4 is predicted to interfere with DNA binding due to an increase in distance between α3 of the dimer mates [13]. This hydrophobic pore is not present in bcPadR structures determined to date.

The genome of hypervirulent *C. difficile* R20291 contains the protein coding sequence for three PadR-like family proteins (cdPadR1, CDR2091_1187, CDR2091_3068). The function of cdPadR1 is of interest due, in part, to its similarity to previously described bcPadRs and LmrR and the response of these transcription regulators to multiple inhibitors. Importantly, differential expression studies have linked cdPadR1 and a homologue from historical *C. difficile* strain 630 (CD630_1154) to regulatory networks that allow *C. difficile* to efficiently respond to environmental changes and, thus, survive within a host. This response is not necessarily due to direct interaction with stressors, but may be part of an overall regulatory cascade. Germination of *C. difficile* strain 630 endospores lead to the differential expression of 92 different transcriptional regulators, ~74 % of which were up-regulated as detected by microarray and validated by qRT-PCR [14]. Included in this list of differentially expressed transcription regulators is the cdPadR1 homologue CD630_1154, which was 2.3-fold up-regulated during germination [14]. This suggests that the expression of one or more of these proteins required to bring an endospore out of dormancy may be regulated by CD630_1154. Another study linked the differential expression of this cdPadR1 homologue to acid and alkali shock, oxygen exposure, and subinhibitory concentrations of metronidazole (Mtz) as detected by microarray analyses in *C. difficile* strain 630 [15].

Herein, we investigated the PadR-s2 protein from *C. difficile* strain R20291, cdPadR1. In this paper, we report the crystallization and X-ray crystal structure of cdPadR1 at 1.9 Å resolution. We also demonstrate cdPadR1 binding to its own gene promoter in a manner conducive to auto-regulation. Additionally, we show that cdPadR1 binds the promoters of three additional regulatory signaling proteins and that a cdPadR1 binding motif is present upstream of 100 genes in *C. difficile* R20291.

Methods
Protein expression and purification
Residues 1-109 of cdPadR1 (locus tag CDR2091_0991) were amplified from gDNA using forward primer

Isom et al. BMC Microbiology (2016) 16:231
Pr3 -EAK (5’-TTTCAGGATCCATGCAGTTAAA TAAAGAAGTGTAAAAAGG-3’ ) and reverse primer Pr4-EAK (5’-TTAACGCTGACGTTAATCCACTCTCC CAAAATTG-3’ ) primers, each of which contained a 5 nucleotide overhang followed by restriction digestion sites for BamHI (forward) or PstI (reverse) for digestion and ligation into the expression vector. cdPadR1 was expressed in Escherichia coli Rosetta™ using the pQE80L (Qiagen) vector system modified to encode a Strep II™-tag on the N-terminus [16]. cdPadR1 was isolated by batch purification over Streptactin SuperFlow Plus resin (Qiagen). All buffers were prepared according to the manufacturers’ guidelines. Cell lysis, column equilibration, and wash buffer contained 50 mM NaH₂PO₄ and 300 mM NaCl guidelines. Cell lysis, column equilibration, and wash buffers were prepared according to the manufacturers’ guidelines. Cell lysis, column equilibration, and wash buffer contained 50 mM NaH₂PO₄ and 300 mM NaCl (pH 8.0 using NaOH). Elution buffer contained 50 mM NaH₂PO₄, 300 mM NaCl, and 2.5 mM d-desethylbiotin (pH 8.0 using NaOH). Subsequent purification of the cdPadR1 dimer was accomplished by size exclusion chromatography in buffer containing 20 mM Tris (pH 8.0 with NaOH) and 150 mM NaCl, using a Superdex 200 Increase 10/300 GL column connected to an ÄKTAs Pure 25 (GE Healthcare). Fractions corresponding to a dimer were concentrated using Amicon® concentration units (Millipore) primed with glycerol and buffer exchanged into 10 mM Tris (pH 8.0) and 100 mM KCl. The molecular weight (MW) was determined by coupling SEC with multi-angle light scattering (MALS) and outputs were analyzed by the ASTRA software (Wyatt Technology).

Crystallization of cdPadR1

Crystals were initially obtained by vapor diffusion using a MCSG Crystallization Suite (Microlytic) (3 M NaCl and 0.1 M HEPES pH 7.5) with a final protein concentration of 1.5 mg mL⁻¹. Crystal growth was optimized at room temperature by hanging drop vapor diffusion with the drops containing 3 μL protein solution (4 mg mL⁻¹ cdPadR1 in 100 mM KCl, 10 mM Tris pH 8.0) and 1 μL reservoir solution (3.1 M NaCl, 100 mM HEPES [pH 7.5]). Crystals were transferred into drops containing an equal volume of 2X reservoir solution and 40 % glycerol for cryoprotection. Crystals in cryosolution were incubated over normal well solution for 5 min before freezing in a liquid nitrogen gas stream for cryogenic data collection.

Data collection and structure determination

X-ray diffraction data were collected using a MARmosaic325 CCD detector at the Stanford Synchrotron Radiation Lightsource (SLAC National Accelerator Laboratory) on beam-line BL14-1. The data were processed with XDS and XSCALE [17]. The XDS output files were converted to .mtz format using CCP4 [18]. The structure of Clostridium thermocellum PadR-like family protein (CtPadR, PDB ID 1XMA) was used as the starting model for molecular replacement using Phaser-MR [19]. The individual coordinates of the preliminary model were generated in AUTOBUILD [20], were refined and rebuilt using the model in COOT [21] and any positions with strong densities outside of the model were accounted for. Structure alignments were performed in COOT and all structure/alignment figures prepared using PyMOL [22]. Residues 1–9, 41, and 107–109 were not modeled due to the absence of electron density. Coordinates have been deposited with the Protein Data Bank (www.rcsb.org) with PDB ID 5DYM. Data collection and refinement statistics are shown in Table 1.

Construction of cdPadR1_W94A

The tryptophan 94 (W94) codon (TGG) of cdPadR1 was converted to alanine (GCG) by overlapping PCR [23]. The sequence of forward and reverse primers used to generate the alanine codon substitutions in cdPadR1 were 5’-GAAACAAAGGAGCATTTATATAAAAAG-3’ and 5’-CTTTTTTAAATATCGTTCTTTGTTTC-3’, respectively. The resulting plasmid was confirmed by sequencing, and the resulting protein variant was overexpressed and purified in the same manner as performed for the native cdPadR1.

Electrophoretic Mobility Shift Assay (EMSA)

Double stranded DNA fragments for EMSA were generated by suspending custom complementary ssDNA (LifeTechnologies) in annealing buffer (10 mM Tris [pH 8.0] and 50 mM NaCl) and heating to 95 °C for 5 min followed by slowly cooling to room temperature. DNA was quantified with the Quant-IT™ Broad Range DNA assay and a Qubit® fluorimeter (Invitrogen). Template dilutions for EMSA stock solutions were dependent on the size of the DNA fragment and ranged from 0.5 μM (100 bp fragment) to 2.5 μM (20 bp fragment). Binding reactions were performed at room temperature. Each reaction mixture contained 20 mM Tris pH 8.0, 120 mM KCl, 12.5 % glycerol, 10 mM MgCl₂, 5 mM DTT, and 125 μg mL⁻¹ heparin. Heparin concentration was increased to 400 μg mL⁻¹ for competition studies. A 1:10 dilution of DNA stock was added to all reactions and a cdPadR1 concentration 2.5–40-fold greater than that of final DNA concentration was added to start the binding reaction. A protein-free control was also included. EMSSAs were performed in 8 % polyacrylamide gels and TB running buffer (89 mM Tris base and 89 mM boric acid) at 200 V and 20–100 mA with run time ranging from 20 min (20 bp fragments) to 30 min (100 bp fragments). Gels were stained with SYBR® Gold Stain (Invitrogen). Image coloration was inverted for ease of viewing. A list of oligonucleotides examined, including location on the genome, sequences, and GC content (%) can be found in Additional file 6: Table S1.
were allowable on either the minus or plus strand of the C. difficile GLAM2 Scan using the motif sites to be on either the plus or minus strand. The GLAM2 was utilized to find a representative Motif Analysis.

Data collection and refinement statistics

| Data collection (SSRL BL14-1)                  |
|-----------------------------------------------|
| Wavelength (Å)                                 | 1.2 |
| Space group                                   | P 4, 2, 2 |
| Unit cell dimensions                           |
| a, b, c (Å)                                    | 92.4, 92.4, 40.4 |
| α, β, γ (%)                                    | 90 |
| Resolution (Å)                                 | 34.4 – 1.9 (2.0 – 1.9) |
| Rmerge (%)                                     | 0.19 (0.38) |
| I/σ(I)                                         | 203 (2.2) |
| Completeness (%)                               | 99 (95) |
| Redundancy (%)                                 | 2.0 (2.0) |
| Wilson B (Å)                                   | 33.1 |

Refinement statistics

| Resolution range (Å)                           | 34.4 – 1.9 (2.0 – 1.9) |
| No. of reflections                             |
| Total                                          | 28484 (2641) |
| Unique                                         | 14302 (1332) |
| Rwork (%)                                      | 0.21 (0.29) |
| Rfree (%)                                      | 0.23 (0.34) |
| No. of atoms                                    |
| Total (non-H)                                  | 842 |
| Water molecules                                 | 34 |
| Mean overall B factors                          |
| Protein                                        | 49.3 |
| Solvent                                        | 43.3 |
| RMSDs                                          |
| Bond length (Å)                                | 0.008 |
| Bond angles (°)                                | 1.03 |
| Ramachandran plot                              | % most favored residues: 96 |
| % outliers                                     | 0 |
| PDB code                                       | SDYM |

Statistics for data in the highest resolution shell are given in parentheses.

Table 1: Data collection and refinement statistics

Motif Analysis

GLAM2 was utilized to find a representative cdPadR1 motif [24]. The sequence surrounding Boxes 1 & 2 (5’-GTACTA TACATTATAGAGTAGTAG-3’) and Boxes 3 & 4 (5’-AGAGTACTAATGTATTATAGTAAAT-3’) were used as input sequences for the GLAM2 analysis. The GLAM2 search was done using the default parameters and allowed the motif sites to be on either the plus or minus strand. The direct GLAM2 output was used as the input for GLAM2 Scan using the C. difficile R20291 genome. Motifs were allowable on either the minus or plus strand of the genome and 200 alignments were allowed. The identified motifs were then mapped onto the C. difficile R20291 genome sequence in Geneious v8 [25]. The motifs were then manually curated to determine whether they were located within an open reading frame, an intergenic promoter region or between convergent genes.

Results and discussion

Crystal structure of recombinant cdPadR1

cdPadR1 shares 100 % amino acid sequence identity with the PadR-like transcription regulator, CD630_1154, in the historical C. difficile strain 630 (Fig. 1), both of which were differentially expressed under conditions of environmental stress [15]. cdPadR1 crystallized in space group P4 2 2 and, following X-ray data collection, the structure was solved by molecular replacement using the PadR family protein from C. thermocellum (CtPadR) as a search model (PDB ID 1XMA). CtPadR and cdPadR1 share 42 % amino acid sequence identity (Fig. 1) and, based on 3D prediction programs [26, 27], were expected to have high structural similarity (RMSD = 1.7 Å). The model was refined to a final crystallographic R-factor of 21.0 % (Rfree = 23.0 %) (Table 1).

One molecule of cdPadR1 was present in the asymmetric unit and consists of an N-terminal winged helix-turn-helix (wHTH) domain (residues 6–80) and a single a-helical C-terminal domain (residues 81–106) (Fig. 2a). This small C-terminal domain places cdPadR1 in the PadR-s2 subfamily of PadR transcriptional regulators described previously [10]. cdPadR1 forms a dimer with a 2-fold crystallographic axis of symmetry (Fig. 2b), similar to the bcPadRs (PDB IDs 4ESB and 4ESF) and LmrR (PDB ID 3F8B), both of which are PadR-s2 family proteins. The dimeric state of cdPadR1 is retained in solution as determined by size exclusion chromatography (Additional file 1: Figure S1). The recognition helices (α3/α3’) are positioned ~34 Å apart (Fig. 2b) consistent with symmetrical binding to two “half-sites” approximately 10 bp in length [28]. Dimerization of cdPadR1 buries approximately 1100 Å2 solvent-accessible surface area (16 %) of the approximately 7000 Å2 total solvent-accessible area per subunit [29]. Residues on helices α1, α2, and α4 that interact to form the cdPadR1 dimer interface are conserved across structural homologues (Fig. 1). The RMSD values for the cdPadR1 structural homologues bcPadR1, bcPadR2, apo-LmrR, LmrR-H33342, and LmrR-daunomycin are 1.6 Å, 1.6 Å, 2.1 Å, 2.9 Å, and 3.3 Å, respectively [27].

The primary helices involved in dimerization are α1 and α4. The amino acid sequence pairwise identities between α1 of cdPadR1 and bcPadR1, bcPadR2, and LmrR are 26 %, 35 %, and 21 %, respectively. The amino acid sequence pairwise identity between α4 of cdPadR1 and bcPadRs (22.2 and 33.3 % for bcPadR1 and bcPadR2,
respectively) is higher than the identity between α4 of cdPadR1 and LmrR (15%). Helix α4 and α′ of cdPadR1 bend toward each other (Fig. 2b) and interact via a coiled-coil, whereas α4 and α′ of LmrR do not display a significant bending towards each other at the C-terminus (Fig. 3a, red). In addition, LmrR contains fewer residues involved in dimerization at the C-terminus of the helix than cdPadR1 and bcPadRs. cdPadR1, like bcPadRs and ctPadR, has a closed dimeric interface, unlike the hydrophobic pore wherein aromatic drug-interaction occurs in LmrR (Fig. 3b). The known structural homologues of cdPadR1 contain a conserved W located within residues 91–96 in the α4 helix region that is predicted to be involved in both dimerization and drug binding [10, 11]. The distance between the conserved W residues in the α4 helix dimer mates for cdPadR1, bcPadR1 (4ESB), apo-LmrR (3F8B), LmrR-H33342 (3F8C), and LmrR-daunomycin (3F8F) was measured using Chimera [30]. Distances were determined from the centroids of the phenol rings (P-P), indole rings (I-I), and indole-to-phenol (I-P) of the conserved α4 W residues. The P-P, I-I, and I-P distances between cdPadR1 W94 and W94′ are 5.4 Å, 9.2 Å, and 7.4 Å, respectively. These distances are similar to those of bcPadR1 (P-P = 5.6 Å, I-I = 9.1 Å, and I-P = 7.4 Å). The P-P distance is ~2 Å greater for the apo-LmrR (P-P = 6.9 Å), LmrR-H33342 (P-P = 7.2 Å), and LmrR-daunomycin (P-P = 7.4 Å) structures when compared to the distance between phenol centroids in cdPadR1. The increased distance between α4 and α′ of LmrR allows for aromatic inhibitor interaction via π-stacking between the W96 and W96′ residues [11]. The lack of a drug-binding pocket in cdPadR1 suggests that any differential expression of the CD630 homologue (CD630_1154) during Mtz exposure would, most likely, be due to a regulatory cascade effect rather than direct interaction of cdPadR1 with Mtz. It was suggested that changes in the orientation of α4 and α′ in a drug-bound state affect the position of the DNA recognition helices, rotating them away from each other [11]. This, presumably, would cause a change in DNA-binding. Previous work revealed that LmrR binds two sites within the lmrCD promoter, one region containing the predicted -10 and -35 sites and the other containing the inverted repeats ATGT/ACAT separated by 10 nucleotides and that this is consistent with a “conserved” binding motif among other PadR-like regulators with an eight nucleotide linker between the inverted repeats ATGT/ACAT [9, 31]. The recognition helices (α3/α3′) are positioned ~34 Å apart in cdPadR1 (Fig. 2b), which is consistent with symmetrical binding to two “half-sites” comprised of inverted repeats ~10 bp apart; it is important to note that this distance does not account for DNA secondary structure. DNA binding behavior was explored for cdPadR1 to determine if it functions similarly to previously studied PadR family transcription regulators and to begin elucidating the regulatory networks of cdPadR1 in hypervirulent C. difficile in vitro.
cdPadR1 binding to its own promoter

A 100 bp region upstream of cdpadR1 (P_{cdpadR1/Pr27}) was used in EMSA assays to determine if cdPadR1 binds its own promoter (Fig. 4a). The presence of five bands with differing mobility indicated that protein-DNA complexes of varying stoichiometry were produced. This may be the result of multiple binding sites and/or higher order oligomerization upon DNA binding (Fig. 4b). Increasing the concentration of cdPadR1 in the reaction resulted in a variation of the migration pattern until an observed saturation point at the slowest mobility compared to other bands was achieved (40-fold cdPadR over DNA or 4 μM cdPadR1, Fig. 4b, far right). cdPadR1 binding to P_{cdpadR1} (Pr27) is consistent with auto-regulation of its own expression.

To further define the binding sites for cdPadR1 binding to P_{cdpadR1}, Palinsight was used to identify inverted repeats within Pr27 characteristic of those bound by transcriptional regulators containing a HTH motif [32–34]. Two sets of inverted repeats (Box 1/2 and 3/4) were identified with a TACT(N_{11-12})AGTA sequence motif (Fig. 4a). A series of smaller dsDNA fragments within the 100 bp P_{cdpadR1} were designed to test the role of these inverted repeats in cdPadR1 binding to P_{cdpadR1} (Fig. 4a). A 64 bp fragment containing both sets of inverted repeats (Pr32) showed four shifts of varying stoichiometry similar to that seen for Pr27 (Fig. 4c). However, full saturation, as seen for Pr27, was not achieved suggesting that additional space on the DNA for higher order oligomerization is needed to see complete shifting to one higher molecular weight complex. When cdPadR1 bound a 61 bp fragment that contained only one set of inverted repeats (Pr31) three shifted complexes were observed (Fig. 4d). This is consistent with the loss of a full binding site and additional space on the DNA for higher order oligomerization as noted for Pr31.

We further narrowed cdPadR1 binding to two small regions of P_{cdpadR1} (Pr68 and Pr122) each containing one set of inverted repeats TACT(N_{11-12})AGTA, shown in EMSA (Fig. 4e and f, respectively). Additionally, a variety of dsDNA fragments representing various sub-regions of the original 100 bp P_{cdpadR1} (Pr27) were examined and, unless the fragment contained the predicted inverted repeats TACT(N_{11-12})AGTA, no binding was observed (Fig. 4g). It was noted that the N_{11-12} spacer region within the inverted repeats was AT rich. To determine whether the AT richness contributes to localized bending of the DNA that facilitates binding we replaced the TTATA in Pr68 with a GCCTG sequence (Pr101). Indeed, significant binding of cdPadR1 to Pr101 was not observed (Additional file 2: Figure S2) suggesting that the AT-rich spacer is important for binding. It should be noted that a fragment containing the AT rich portion
but lacking the intact TACT/AGTA (Pr42) was not bound by cdPadR1 (Fig. 4a and g). This indicates that the AT rich sequence is not the direct binding site for cdPadR1. Additionally, varying the length of the spacer between the TACT/AGTA inverted repeat in Pr68 did not interfere with binding (Additional file 2: Figure S2) suggesting that flexibility of the DNA region between the inverted repeat rather than the length is more important for cdPadR1 binding.

To summarize, cdPadR1 binding to P_{cdpadR1} is dependent upon a TACT/AGTA inverted repeat sequence. Two such sequences are present in the 100 bp P_{cdpadR1}-investigated in this study. These two inverted repeats are responsible for two sequence-specific interactions between cdPadR1 and its promoter that can account for two shifted complexes. Additional shifted complexes may be the result of higher order oligomerization of cdPadR1 once bound to DNA or a decrease in the constraints on sequence specificity. Although constraints on the spacing between the TACT/AGTA inverted repeats do not appear to be tight, there does appear to be a requirement for AT richness within the spacer. The placement of the inverted repeat within P_{cdpadR1} is consistent with auto regulation. cdPadR1 and the cdPadR1 homologue CD630_1154 both contain TACT/AGTA with an 11 nucleotide spacer 25 bp upstream of the open reading frame (ORF). Additionally, both promoter regions contain TACT(N_{12})AGTA 52 bp upstream of their respective ORFs and overlapping the predicted -35/-10 promoter region, which suggests a similar binding function for each of these genes to their respective promoters.

cdPadR1 binds other gene promoters with the cdPadR1 motif
The dsDNA fragments containing TACT(N_{11,12})AGTA from P_{cdpadR1} were analyzed for conserved binding motif using GLAM2 [24]. GLAM2 was advantageous over MEME because it allows for spacing/gaps in motif
prediction since spacing between the inverted repeats was not critical for binding. The best motif was 21 bp in length with the sequence GTACTAT(N2)ATTATA(N) AGTA and was designated cdPadR1 motif (Fig. 5a). GLAM2Scan results indicated the presence of 200 potential motif matches in the _C. difficile_ strain R20291 genome with scores ranging from 13.6–18.7, not including the P_{padR1} sequences used for analysis (Additional file 3: Table S2). Approximately half of these motifs are either situated between two convergent genes or are located within open reading frames (ORFs). Of those that are located upstream of genes, approximately 6 % are upstream of other transcription regulators and other regulatory proteins, such as two-component response regulators, while another ~7 % are upstream of genes involved in transport/efflux and sporulation. The genes predicted to be involved in transport/efflux are the ABC transporter ATP-binding proteins CDR20291_0159, _0296, _0551, _0553, and _3203 (Additional file 4: Table S3). Two genes predicted to be involved in sporulation also contain the cdPadR1 binding motif upstream of the transcription start site, a spore maturation protein (CDR20291_3377) and a spore coat assembly protein (CDR20291_0316) (Additional file 4: Table S3). Over 50 % of the predicted binding motifs were indicated to be either upstream genes of “hypothetical proteins”, within open reading frame, or between convergent genes. Exemplar promoters from this list were selected for analysis using EMSA to determine binding of cdPadR1 to these promoter fragments in vitro (Fig. 5b and c). A 30 bp and 100 bp dsDNA fragment was selected for each promoter region and contained at least one predicted cdPadR1 motif (Fig. 5b and c). Pr132 and Pr133 contain the cdPadR1 motif located 45 base pairs upstream of CDR20291_2322 (IclR family transcription regulator CDS) (cdP_{2322}). Pr135 and Pr136 contain the cdPadR1 motif located 116 base pairs upstream of CDR20291_1882 (two-component system response regulator CDS) (cdP_{1882}). Pr137 and Pr138 contain the cdPadR1 motif located 25 base pairs upstream of CDR20291_1590 (ArsR family transcriptional regulator CDS) (cdP_{1590}). cdPadR1 bound all of the selected promoters in vitro (Fig. 5c). The 30 bp promoters (Pr132, Pr135, Pr137) yielded two discrete bands. However, this phenomenon has also been observed on occasion for the short dsDNA fragment containing one set of
inverted repeats from P\textsubscript{cdpadR1} (Pr68, Additional file 5: Figure S3) and is not likely to represent multiple binding events to a small dsDNA fragment [35]. This binding pattern may be attributable to the presence of small amounts of ssDNA, portions of the dsDNA with secondary structure, or conformational changes in the DNA upon binding in a small subset of the complexes which are more pronounced in shorter dsDNA fragments [35].

Gene regulatory networks play an integral role in the physiology of microorganisms and their response to ever changing environments [36, 37]. The binding of cdPadR1 to the promoters of genes encoding transcription regulators and a DNA-binding response regulator, part of a two-component signal transduction system, suggests it may play a role in a gene regulatory network in C. difficile. The cdPadR1 motif overlaps the predicted -10 region of cdP\textsubscript{1590} and cdP\textsubscript{2322} [38]. This positioning of a regulatory binding site overlapping the -10 region is consistent with repression via abrogation of the Sigma factor. In cdP\textsubscript{1882} the cdPadR1 motif is located approximately 30 bp upstream of the predicted -35 region [38]. Positioning of a regulatory binding site upstream of the -35/-10 core promoter elements is typically consistent with a role in activation of the promoter [39]. While additional studies are necessary to determine the biological role of cdPadR1 in activation or repression of these promoters, it is notable that cdPadR1 is able to bind these promoters and likely participates in a regulatory cascade in response to undetermined stimuli.

\textbf{cdPadR1 binds other promoter regions}

Additional promoters from the cdpadR1 genomic neighborhood were chosen to test for cdPadR1 binding based on gene expression studies. A promoter for a nitric oxide reductase (norV, CDR20291_0994) and a Spo0B-associated GTP-binding protein were selected. Nitric oxide reductase has been linked to pathogenesis in other microorganisms [40] and was 2-fold down regulated, along with cdPadR1 when compared to the historical C. difficile strain 630 [41]. Another representative promoter for EMSA study from within cdPadR1 genome neighborhood is upstream of a gene encoding a Spo0B-associated GTP-binding protein (obg, CDR20291_1001) whose homologue was 2-fold down regulated following pig loop infection with the historical strain C. difficile 630 [42]. cdPadR1 bound P\textsubscript{norV} and P\textsubscript{obg} in vitro (Fig. 6d and e, respectively). The migration patterns for P\textsubscript{norV} and P\textsubscript{obg} differ from that of P\textsubscript{cdpadR1} (Fig. 6c). For all promoters examined, slower migrating complexes appeared at increasing protein concentrations, which suggests that cdPadR1 binds to multiple sites in the upstream region of the gene. However, the complexes formed when cdPadR1 is incubated with promoters other than its own are smaller and it appears that a level of saturation, wherein only one large complex is formed, is not reached as it is for P\textsubscript{cdpadR1}. It is well understood that transcription regulators bind a relatively limited set of DNA sequences [43], a concept that we explored for cdPadR1 and P\textsubscript{cdPadR1} (Fig. 4), as well as a predicted binding motif (cdPadR1 motif, Fig. 5). Both P\textsubscript{norV} and P\textsubscript{obg} have only one half of the inverted repeat within the cdPadR1 motif (Fig. 6a and b). However, it is unclear whether only one half-site is sufficient to initiate binding to these promoters or if perhaps the binding is non-specific and related to local DNA structure or AT content. Therefore, we examined binding specificity

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig6}
\caption{EMSAs of cdPadR1 binding the cdpadR1 promoter (P\textsubscript{cdpadR1}), a norV promoter (P\textsubscript{norV}), and b obg promoter (P\textsubscript{obg}). The predicted -10 and -35 sites are indicated in blue boxes above the sequence. Orange arrows indicate the inverted repeats TACT/AGTA. c e Final P\textsubscript{cdpadR1}, P\textsubscript{norV}, and P\textsubscript{obg} (100 bp) concentration in the reaction was 0.1 \textmu{}M. cdPadR1 was 5, 10, 20, and 40-fold excess over dsDNA. The minus (-) lane contains dsDNA and no cdPadR1. Shifted DNA-protein complexes are annotated with a black arrow and unbound dsDNA migration is marked with a red arrow. f For this EMSA, the final heparin concentration was 4-fold of the standard EMSA concentration used throughout this research. The + lane contains 40-fold excess cdPadR1 over dsDNA.}
\end{figure}
using increased amounts of heparin as a competitor for cdPadR1 binding (Fig. 6f). When a 4-fold higher concentration of heparin was present in the binding reaction of cdPadR1 to P_{norV} or P_{obs} a shifted complex was no longer detected at 40-fold protein over dsDNA (Fig. 6f). Under the same conditions, cdPadR1 still bound its own promoter, though the larger complexes were no longer detected. That cdPadR1, a small HTH DNA binding protein, would bind other 100 bp predicted promoter regions non-specifically could be explained using the theoretical model termed one-dimensional diffusion, or “sliding”. During one-dimensional diffusion, the transcriptional regulator searches for specific binding sites along the DNA remaining in contact with the DNA due to non-specific interactions [44–46]. It is, therefore, likely that a more specific level of binding requires the full cdPadR1 motif. So, while cdPadR1 does, in fact, bind P_{norV} and other 100 bp AT-rich promoters in vitro (Additional file 6: Table S1), no conclusions can be made regarding the regulation of this or any other promoters tested based on EMSA alone. Coupled with the recent expression studies, however, in vitro binding assays suggest that further study into the regulation of expression of these genes, especially norV, is warranted.

Role of the conserved W residue in cdPadR1 DNA binding
It was suggested previously that the conformational changes elicited by drug binding between α4′/α4′ could affect DNA binding and that a conserved tryptophan (W) in α4 was directly involved in drug binding; an indirect role of this W residue was indicated in DNA binding [11]. We examined the effect of this conserved W at residue 94 (W94) in cdPadR1 on DNA binding in vitro (Fig. 7). When W94 is replaced with an alanine (cdPadR1W94A), the majority of binding along with the higher order complexes observed for cdPadR1WT binding to P_{cdpadR1} are lost (Fig. 7). Dimerization was not as detected by size exclusion chromatography coupled with multi-angle light scattering detection (SEC-MALS, Additional file 1: Figure S1). These results suggest that, while the conserved W does not affect dimerization, it does inhibit DNA binding in vitro in a way that is not entirely clear while further supporting a role of the conserved W in DNA binding. The suggested mechanism by Madoori et al wherein the DNA binding helices of LmrR putatively rotate away from each other when the effector-binding/oligomerization domain is perturbed at the conserved W residue is further supported as the mechanism of lowered DNA binding affinity by the results presented here.

Conclusion
We have determined the 1.9 Å resolution crystal structure of cdPadR1, which revealed that it is in the PadR-s2 subfamily of PadR transcriptional regulators with other structurally and functionally characterized PadR-like regulators from B. cereus (bcPadR1 and bcPadR2) and L. lactis (LmrR). In vitro protein-DNA binding experiments demonstrate that cdPadR1 binds a region comprised of the inverted repeats TACT/AGTA and an AT-rich core, GTACTAT(N)_2 ATTATA(N)AGTA, within its own promoter. These predicted binding sites are present in the cdPadR1 homologue CD630_1154, suggesting that these transcription regulators are functional homologues as well. cdPadR1 appears to be part of a hierarchical gene regulatory network in C. difficile. Furthermore, cdPadR1 non-specifically associates with longer DNA fragments that may facilitate promoter and motif searching. Mutation of the highly conserved W in the α4 helical region, which is predicted to be involved in multidrug recognition and dimerization in LmrR, resulted in alterations of cdPadR1 binding to the predicted binding motif, potentially due to tighter constraints on spacing of the inverted repeats as well as a loss of higher order oligomerization. Complementary in vivo studies of cdPadR1 will allow for a better understanding of its regulatory network.

Additional files

Figure S1. (A) Chromatogram from size exclusion chromatography (SEC) run performed on a Superdex 200 Increase 10/300 GL column connected to an AKTA Pure 25 (GE Healthcare). The black line represents the calibration standard mix with molecular weights of standards labeled. The blue and red lines indicate the elution profile for cdPadR1WT and cdPadR1W94A, respectively (13.5 kDa monomer size for both). (B) Molar mass versus elution time of cdPadR1WT and cdPadR1W94A from SEC (as described) coupled with multi-angle light scattering (MALS) detection. Red lines indicate MALS signal (LS) and green lines indicate UV detection. cdPadR1WT and cdPadR1W94A both dimer with molecular weights (MW) of approximately 30 and 27 kDa, respectively (monomeric cdPadR1 MW is 13.5 kDa). (PPTX 222 kb)

Figure S2. EMSA of cdPadR1 binding different fragments of P_{cdpadR1}, that contain the inverted repeats TACT/AGTA with 4 bp overhang on the 5′ and 3′ end of inverted repeats. Each dsDNA contains a different number of nucleotides between TACT/AGTA (addition of a central alanine). EMSAs were conducted as described for 100 bp and smaller dsDNA fragments in the materials and methods. A
complete list of nucleotides tested is available in the table below the EMDA gel. Inverted repeats are underlined. For Pr101, the AT rich region that was mutated is indicated in bold. The - lane contains DNA only and the + lane contains 10-fold cdPadR1 in excess over DNA. (PPTX 256 kb)

Additional file 3: Table S2. List of GLAM2Scan [24] results for scan of the Clostridium difficile R20291 genome for cdPadR1 motif (Fig. 5a). The.xlsx file includes the locus ID for the gene(s) nearest the motif (column A); the gene annotation, if applicable, or the indication that this motif is within an open reading frame (ORF) or between two convergent genes (column B); the genomic placement (start, column C and end, column D); indication of whether the motif is on the leading (+) or lagging (-) strand (column E); the motif sequence generated by GLAM2Scan (column F); and the score generated by GLAM2Scan that indicates how well the sequence fits the cdPadR1 motif (column G). (XLSX 49 kb)

Additional file 4: Table S3. List of GLAM2Scan [24] results for scan of the Clostridium difficile R20291 genome for cdPadR1 motif (Fig. 5a) for promoters upstream of genes predicted to be involved in transport/exflux and sporulation. The .xlsx file includes the locus ID for the gene(s) nearest the motif (column A); the gene annotation (column B); the genomic placement (start, column C and end, column D); indication of whether the motif is on the leading (+) or lagging (-) strand (column E); the motif sequence generated by GLAM2Scan (column F); and the score generated by GLAM2Scan that indicates how well the sequence fits the cdPadR1 motif (column G). (XLSX 39 kb)

Additional file 5: Figure S3. EMDA of cdPadR1 binding the 21 bp fragment (Pr68. Additional file 4: Table S1) that contains the inverted repeats TACT/AGTA with 11 nucleotides in between from within its own promoter. Protein-free controls are indicated with a minus sign (−). 21 bp PA20291_1590 DNA (0.25 μM) was used in a reaction with increasing concentrations of cdPadR1 (2.5, 5.0, and 100.0 μM). (PPTX 3693 kb)

Additional file 6: Table S1. Full list of oligonucleotides that were annealed and used in EMSA studies with cdPadR1. The .xlsx file includes the arbitrary number assigned to each oligonucleotide (column A); the genomic placement of the minimum nucleotide (column B); length in base pairs (bp, column C); indication of binding (+) or no binding detected (−) (column D); locus tag associated with the gene downstream of the oligonucleotide (column E); the annotated gene downstream of the oligonucleotide (column F); and the oligonucleotide sequence 5′ to 3′ (column G). (XLSX 50 kb)

Abbreviations
A: Alanine; Å: Ångström; Abbreviations
BC4206; norV
M: Molar; MALS: Multi-angle light scattering; MCSG: Midwest Center for Structural chloride; L: Liter; LmrR: Lactococcal multidrug resistant transcription regulator; HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; HTH: Helix-turn-motifs version 2; GTP: Guanosine triphosphate; H33342: Hoeschst 33342; DTT: Dithiothreitol; EMSA: Electrophoretic mobility shift assay; DNA: Deoxyribonucleic acid; dsDNA: Double stranded deoxyribonucleic acid; DTI: Dithiothreitol; EMDA: Electrophoretic mobility shift assay; gDNA: Genomic deoxyribonucleic acid; GLAM2: Gapped local alignment of motifs version 2; GTP: Guanosine triphosphate; H33342; Hoeschst 33342; HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; HTH: Helix-turn-helix; ID: Identification; I1: Indole-indole; I1P: Indole-phenol; KCl: Potassium chloride; L: Litter; LmrR: Lactococcal multidrug resistant transcription regulator; M: Molar; MAL: Multi-angle light scattering; MCGS: Midwest Center for Structural Genomics; MEME: Multiple Em for Motif Elicitation; MgCl2: Magnesium chloride; mM: Millimolar; MW: Molecular weight; N: Any nucleotide; NaCl: Sodium chloride; NaH2PO4: Sodium phosphate; NaOH: Sodium hydroxide; mM: Millimolar; MW: Molecular weight; N: Any nucleotide; NaCl: Sodium chloride; NaH2PO4: Sodium phosphate; NaOH: Sodium hydroxide; mM: Millimolar; MW: Molecular weight; N: Any nucleotide; NaCl: Sodium chloride; NaH2PO4: Sodium phosphate; NaOH: Sodium hydroxide; mM: Millimolar; MW: Molecular weight; N: Any nucleotide; NaCl: Sodium chloride; NaH2PO4: Sodium phosphate; NaOH: Sodium hydroxide; mM: Millimolar; MW: Molecular weight; N: Any nucleotide; NaCl: Sodium chloride; NaH2PO4: Sodium phosphate; NaOH: Sodium hydroxide; mM: Millimolar; MW: Molecular weight; N: Any nucleotide; NaCl: Sodium chloride; NaH2PO4: Sodium phosphate; NaOH: Sodium hydroxide; mM: Millimolar; MW: Molecular weight; N: Any nucleotide; NaCl: Sodium chloride; NaH2PO4: Sodium phosphate; NaOH: Sodium hydroxide; mL: microliter; μM: micromolar

Acknowledgements
We would like to thank Ana Gonzalez for help with the data collection during the RapıData 2015 course. We also thank the 2015 RapıData Course for access to the Stanford Synchrotron Radiation Lightsource (SSRL) beamline (BL-14) for diffraction data collection. SSR1 is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). Additionally, we would also like to thank Dr. Jimmy Ballard for Clostridium difficile gDNA. We are also grateful to Neda Hessami (OU Michael F. Price Institute for Structural Biology) for large-scale sample preparations and helpful discussions. We thank Fares Najer, Jamie Sykes, Samantha Powell, Bing Wang and Skyley Heddon (University of Oklahoma); and Steve Almo, James Love, and Vern Schramm (Albent Einstein College of Medicine) for thoughtful discussions. We are grateful to Dr. Eliza Ruben for sample processing assistance in the OU Protein Production Core facility, which is funded by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20GM103640 to Ann H. West. Initial crystallization trails were carried out in the OU Macromolecular Crystallography Laboratory which is partially funded by National Science Foundation Major Research Instrumentation award 092269.

Funding
This study was supported by private funding from The Price Family Foundation (New York, NY). The funding organization had no involvement in study design, data analysis or interpretation.

Availability of data and materials
All expression plasmids utilized in this study are available upon request. Primer/Oligonucleotide sequences are available in Additional file 4: Table S1. Atomic coordinates were submitted to the Protein Data Bank under accession number SDYM (http://www.rcsb.org/pdb/home/home.do) and will be publically available upon publication. All additional data generated or analyzed during this study are included in this published article [and its supplementary information files].

Authors’ contributions
O is responsible for all protein purification and EMSA experiments as well as X-ray data collection. SM and LT assisted CI in structure determination. AW, GR and EA are the project PI’s that were involved in study design and oversight. All authors participated in manuscript preparation. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not Applicable.

Author details
1 Price Family Foundation Institute of Structural Biology and Department of Microbiology and Plant Biology, University of Oklahoma, 770 Van Vleet Oval, Norman, OK 73019, USA. 2 Price Family Foundation Institute of Structural Biology and Department of Chemistry and Biochemistry, University of Oklahoma, 101 Stephenson Parkway, Norman, OK 73019, USA.

Received: 4 August 2016 Accepted: 24 September 2016

Published online: 04 October 2016

References
1. Tickler IA, Goering RV, Whitmore JD, Lynn AN, Persing DH, Tenover FC, Healthcare Associated Infection C. Strain types and antimicrobial resistance patterns of Clostridium difficile isolates from the United States, 2011 to 2013. Antimicrob Agents Chemother. 2014;58(7):4214–8.
2. Rupnik M, Wilcox MH, Gerding DN. Clostridium difficile infection: new developments in epidemiology and pathogenesis. Nat Rev Microbiol. 2009;7(5):326–36.
