BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

McbR/YncC: Implications for the Mechanism of Ligand and DNA Binding by a Bacterial GntR Transcriptional Regulator Involved in Biofilm Formation

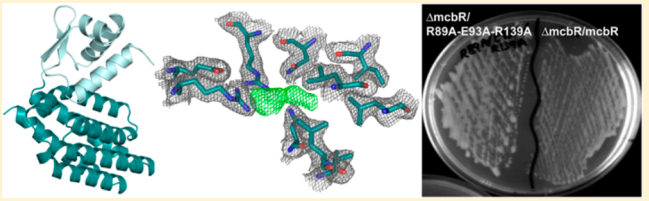
Dana M. Lord,[†,‡] Ayse Uzgoren Baran,[§,⊥] Valerie W. C. Soo,[∥] Thomas K. Wood,[∥] Wolfgang Peti,[§] and Rebecca Page*[,†]

[†]Department of Molecular Biology, Cell Biology and Biochemistry, [‡]Graduate Program in Molecular Pharmacology and Physiology, and [§]Department of Molecular Pharmacology, Physiology and Biotechnology & Chemistry, Brown University, Providence, Rhode Island 02903, United States

[∥]Departments of Chemical Engineering & Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania 16802, United States

Supporting Information

ABSTRACT: MqsR-controlled colanic acid and biofilm regulator (McbR, also known as YncC) is the protein product of a highly induced gene in early *Escherichia coli* biofilm development and has been regarded as an attractive target for blocking biofilm formation. This protein acts as a repressor for genes involved in exopolysaccharide production and an activator for genes involved in stress response. To better understand the role of McbR in governing the switch from exponential growth to the biofilm state, we determined the crystal structure of McbR to 2.1 Å. The structure reveals McbR to be a member of the FadR C-terminal domain (FCD) family of the GntR superfamily of transcriptional regulators (this family was named after the first identified member, GntR, a transcriptional repressor of the gluconate operon of *Bacillus subtilis*). Previous to this study, only six of the predicted 2800 members of this family had been structurally characterized. Here, we identify the residues that constitute the McbR effector and DNA binding sites. In addition, comparison of McbR with other members of the FCD domain family shows that this family of proteins adopts highly distinct oligomerization interfaces, which has implications for DNA binding and regulation.



Biofilms are complex communities of bacteria that are encased in an extracellular matrix and adhere to almost any surface. Due to properties of the biofilm, these bacterial communities are extremely tolerant to antibiotics and are often able to evade host defenses.[1] Furthermore, it is estimated that 60−80% of human infections are caused by biofilms, explaining why much research is focused on elucidating the genetic basis of biofilm formation and proliferation.[2,3] One regulator of biofilm formation is McbR/YncC (hereafter referred to as McbR), a transcription factor predicted to belong to the GntR family of DNA binding proteins. In *Escherichia coli*, deletion of *mcbR* results in the overproduction of colanic acid,[4] a constituent of the biofilm exopolysaccharide (EPS) matrix composed of glucose, galactose, fucose, and glucuronic acid in the ratio 1:2:2:1.[5] As a consequence, *mcbR* deletion results in a mucoidy phenotype and a reduction in biofilm formation. In *E. coli*, gene array studies coupled with electrophoretic mobility shift assays (EMSAs) showed that McbR binds the *ybiM* promoter, a gene encoding a putative periplasmic protein whose function is currently unknown.[4] A subsequent study using DNA footprinting experiments showed that McbR from *Salmonella typhimurium* and *E. coli* binds the *yciG* promoters from both species.[6]

The GntR superfamily (Pfam PF00392), to which McbR belongs, is one of the largest families of transcriptional regulators, with more than 8500 members (Figure 1A).[7] Members of this family contain an N-terminal DNA binding winged helix-turn-helix (wHTH) domain and a C-terminal effector binding/oligomerization domain. In contrast to the wHTH domain, which is structurally conserved in the GntR family, the C-terminal domain is highly variable. Detailed bioinformatics studies have led to the definition of at least 7 families (AraR, DevA, FCD, HutC, MocR, PlmA, and YtrA), which are classified by the effector binding domain topology and secondary structure (Figure 1A).[7] The majority of GntR regulators belong to the FadR C-terminal domain family (FCD, Pfam PF07729; Figure 1A). The effector molecule that regulates the activity of GntR transcriptional regulators is often a product/substrate in the metabolic pathway that the particular GntR transcription factor controls. However, there are examples where this is not the case, and difficulties in identifying the endogenous ligands for this family have limited

Received: July 15, 2014
Revised: October 23, 2014
Published: October 23, 2014

dx.doi.org/10.1021/bi500871a | *Biochemistry* 2014, 53, 7223−7231

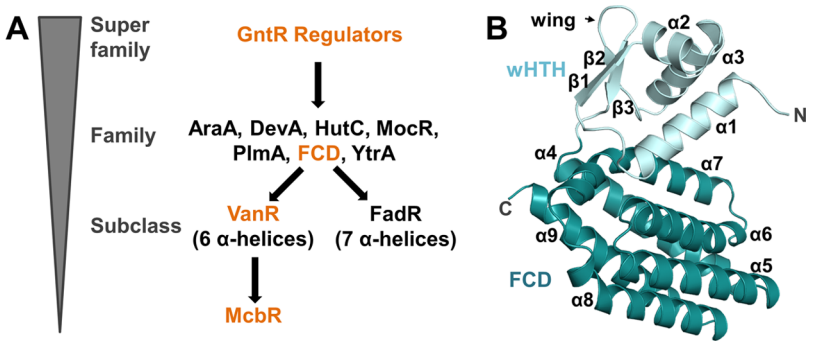


Figure 1. McbR classification and structure. (A) Flowchart illustrating the classification of McbR within the GntR superfamily. (B) McbR monomer with all secondary structural elements annotated. The N-terminal winged helix-turn-helix (wHTH) domain is shown in light blue, and the C-terminal FCD domain is shown in teal; the wing loop of the wHTH domain is labeled. McbR residues 10−220 were observed in the electron density maps for subunit A.

our understanding of how these regulators function *in vivo*.[7] What is known is that effector binding in the C-terminal effector binding domains alter, via a poorly understood structural mechanism(s), the conformations and/or relative orientations of the N-terminal wHTH domains. This, in turn, inhibits DNA binding.[7] This limited understanding is due to the dearth of structural data available for the GntR superfamily, especially those of the FCD family. Here, we describe the structure of McbR from *E. coli* to 2.1 Å resolution. We show that McbR belongs to the FCD family of transcriptional regulators and identify the residues that mediate DNA binding. We also identify the residues that constitute its effector binding site, which are highly conserved in *mcbR* homologs. Finally, a comparison of the currently available structures of FCD transcriptional regulators reveals different oligomerization interfaces at the wHTH domains,[8−10] suggesting that this family of proteins undergoes distinct conformational rearrangements upon ligand binding.

MATERIALS AND METHODS

Protein Expression and Purification. Two constructs of wild-type McbR ($McbR_{1-221}$ (full-length), residues 1−221; $McbR_{10-221}$, residues 10−221) were subcloned into the pRP1B bacterial expression vector, which contains an N-terminal His_6-tag and Tobacco Etch Virus (TEV) cleavage site;[11] both constructs were sequenced prior to subsequent experiments. pRP1B-$McbR_{1-221}$ variants (single-mutant variants: Arg34Ala, Lys38Ala, Thr49Ala, Arg52Ala, Gln70Ala; double-mutant variant: Glu93Ser/Arg139Phe; triple-mutant variant: Arg89Ala/Glu93Ala/Arg139Ala) were generated using the QuikChange Mutagenesis Kit (Agilent Technologies) using the manufacturer's protocols; all constructs were verified by sequencing.

WT McbR and McbR variants were expressed in *E. coli* BL21-Gold (DE3) cells (Agilent). Cells were grown at 37 °C (250 rpm) to an OD_{600} of ∼0.9, at which point the cells were transferred to 4 °C for 1 h. The cells were induced with 0.5 mM IPTG and grown overnight at 18 °C (250 rpm). The cells were then harvested by centrifugation at 6000*g*. Selenomethionine (SeMet)-labeled $McbR_{10-221}$ was produced using identical protocols, with the exception that the cells were grown in minimal medium supplemented with vitamins, metals, and amino acids (with selenomethionine substituted for methionine).[12]

For purification, cell pellets of either WT McbR or McbR variants were resuspended in lysis buffer (50 mM Tris, pH 8.0, 500 mM NaCl, 0.1% Triton X-100, 5 mM imidazole, complete tabs-EDTA free [Roche]) and lysed by high-pressure homogenization (C3 Emulsiflex; Avestin). Following centrifugation (45 000*g*, 45 min, 4 °C), the supernatant was applied to a HisTrap HP column (GE Healthcare), and McbR was eluted using a 5−500 mM imidazole gradient. McbR was then incubated overnight with TEV protease (50 mM Tris, pH 8.0, 500 mM NaCl, 4 °C). The following day, McbR was further purified using Ni-NTA (Qiagen) to isolate the cleaved protein from the TEV protease (itself His_6-tagged) and the cleaved His_6-tag. After concentration, McbR was purified in a final step using size-exclusion chromatography (SEC; Superdex 200 26/60; 20 mM Tris, pH 7.8, 100 mM NaCl, 0.5 mM TCEP). To determine the oligomerization state of McbR, the elution volume was compared to that of MW weight standards (Bio-Rad).

Crystallization, Data Collection, and Processing. SeMet $McbR_{10-221}$ was concentrated to 8−10 mg/mL, incubated for 1 h with glycerol (10% (v/v), final concentration), and used immediately for crystallization trials. Microcrystals of SeMet $McbR_{10-221}$ were obtained in 2 M sodium malonate pH 7.0 (sitting drop vapor diffusion; 25 °C) and used as seed solution to produce crystals in the same conditions suitable for data collection. The crystals were cryoprotected in mother liquor containing 20% (v/v) MPD and immediately frozen in liquid nitrogen. Data for $McbR_{10-221}$ were collected at the National Synchrotron Light Source, beamline X25, using a Pilatus 6 M detector (Dectris). Anomalous data was collected from a single crystal and phased using single anomalous dispersion (SAD), collecting data at 0.93 Å. Data were processed and scaled using HKL2000.[13] The asymmetric unit contains two protein molecules. The anomalous data was phased using HKL2MAP[14] (ShelxC/D/E);[15−17] 14 of the 16 expected selenium sites were identified. Approximately 90% of the structure was built automatically using ARP/wARP.[18] Model building and refinement of SeMet McbR was carried out using a high-resolution data set (2.1 Å) collected at 0.9793 Å. Iterative model building and refinement were performed using COOT[19] and Phenix.[20] The final model was refined with Phenix using TLS. Molprobity was used for model validation.[21] Analysis of the dimerization interface was performed using the Protein Interaction Calculator,[22] with solvent-accessible surface areas calculated using Naccess.[23] Cavity volumes were calculated using POCOSA.[24] Data collection and structure refinement statistics are reported in Table 1.

dx.doi.org/10.1021/bi500871a | *Biochemistry* 2014, 53, 7223−7231

Table 1. Crystallographic Data Collection and Refinement Statistics

Crystal Data		
space group	$P6_3$	
no. McbR/ASU	2	
a, c (Å)	a = 107.6, c = 72.7	
Data Collection		
wavelength (Å)	0.9793	0.9300
unique reflections	28120	50871
resolution (Å)[a]	50.0−2.1 (2.14−2.10)	50.0−2.15 (2.19−2.15)
mean I/σ	18.1 (3.0)	18.4 (2.5)
completeness (%)	100.0 (100.0)	99.3 (98.7)
redundancy	9.8 (10.0)	5.1 (5.2)
R_{merge} (%)[c]	8.3 (55.8)	6.1 (45.4)
Refinement		
R_{work} (%)	16.8	
R_{free} (%)[b]	20.8	
no. non-hydrogen atoms	3214	
no. water molecules	163	
Ave. B-factor (Å^2)		
protein	43.6	
waters	43.4	
rmsd bond length (Å)	0.008	
rmsd bond angle (deg)	1.003	
Ramachandran Plot		
favored (%)	99.8	
allowed (%)	100.0	
disallowed (%)	0	
PDB code	4P9F	

[a]Highest resolution shell data are shown in parentheses. [b]5% of the reflections used for R_{free}. [c]$R_{merge} = \sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$ where $I_i(hkl)$ is the ith observation of a symmetry equivalent reflection hkl. Reported values for the 0.93 Å data set are for unmerged Friedel pairs.

Electrophoretic Mobility Shift Assay. The *E. coli yciG* promoter ($PyciG_{ECO}$) was used for DNA binding studies. Following synthesis of the individual oligonucleotides (IDT Technologies; each oligonucleotide includes a 3′ biotin label; Table 2), the complementary oligonucleotides were combined, heated at 95 °C, and then cooled at 1 °C/min to a final temperature of 25 °C. For EMSA experiments, 1 pmol of protein was added to the biotin-labeled DNA ($PyciG_{ECO}$, 100 fmol). All reactions were carried out in binding buffer (10 mM Tris, pH 7.5, 50 mM KCl, 1 mM DTT) in the presence of a poly(dI−dC) DNA probe (50 ng/μL) to prevent nonspecific binding. For the unlabeled competitor EMSA control, a 200-fold excess of unlabeled $PyciG_{ECO}$ DNA was added. All binding reactions were incubated at room temperature for 20 min. Samples were then loaded onto a 6% DNA retardation gel (Invitrogen) and subjected to electrophoresis at 4 °C for 75 min at 100 V in 0.5-fold TBE buffer (45 mM Tris, pH 8.3, 45 mM Boric acid, 1 mM EDTA). The DNA was transferred to a nylon membrane at 390 mA for 30 min, followed by UV cross-linking at 302 nm by placing the membrane face-down on a UV illuminator for 15 min. Chemiluminescence was performed using the LightShift Chemiluminescent EMSA Kit (Pierce), and the samples were detected using a CCD imager (Typhoon 9410 Imager).

Mucoidy Assay. WT *mcbR* and two mutated *mcbR* variants (*mcbR*-E93S-R139F and *mcbR*-R89A-E93A-R139A) were subcloned from pRP1B-*mcbR*, pRP1B-*mcbR*-E93S-R139F, and pRP1B-*mcbR*-R89A-E93A-R139A into the *Kpn*I and *Sac*I sites of pBS(Kan)[25] using primers in Table 3 so that their expression is under control of a *lac* promoter (instead of a T7 promoter in the former plasmids). The resulting plasmids are pBS(Kan)-*mcbR*, pBS(Kan)-*mcbR*-E93S-R139F, and pBS(Kan)-*mcbR*-R89A-E93A-R139A. After verifying these plasmids by DNA sequencing, they were introduced into *E. coli* MG1655 Δ*mcbR* Δ*Km*[4] via electroporation, and the transformed clones were plated on LB agar supplemented with 50 μg/mL kanamycin and 0.2% (w/v) glucose. At least three independent colonies were streaked on LB agar supplemented with 50 μg/mL kanamycin and 1 mM IPTG to test the mucoidy of each strain. Cells were incubated at 37 °C for 12 h.

RESULTS AND DISCUSSION

McbR Is a Member of the VanR Subfamily of GntR Transcriptional Regulators. Two constructs of McbR were screened for their ability to form diffraction-quality crystals: $McbR_{1-221}$ and $McbR_{10-221}$. The latter is missing the first 9 amino acids, which were predicted to be disordered (PSIPRED,[26] IUPRED[27,28]). Only $McbR_{10-221}$ formed crystals suitable for structure determination and is referred to hereafter as McbR. The crystal structure of McbR was determined by single-wavelength anomalous dispersion (SAD) using SeMet-labeled protein, and the atomic model was refined to 2.1 Å resolution (Table 1 and Figures 1B and 2A). Two molecules of McbR are present in the asymmetric unit and are related by a nearly perfect 2-fold axis (179.8°; superposition using the C-terminal FCD domain; Figure 2A). This is consistent with the observation that McbR is predominantly a dimer in solution (Figure 2B). McbR, like other members of the GntR family, consists of an N-terminal wHTH domain (residues 10−76; residues 38−46 in the second subunit were not modeled due to a lack of clear electron density) and a C-terminal all α-helical effector binding domain (residues 77−219; Figure 1B). The wHTH domain is composed of three α-helices (α1−α3) and three β-strands (β1−β3), which form a small β-sheet that is a defining characteristic of the wHTH fold. The C-terminal domain is composed of six α-helices (α4−α9). The secondary structure elements and topology of the C-terminal domain places McbR in the FadR C-terminal domain (FCD) family of GntR transcriptional regulators (Figure 1A).

McbR Dimerization Interface Is Extensive and Includes Both the N- and C-Terminal Domains. The FCD family has ∼2800 members from more than 400 distinct species from archaea to eukaryota.[10] A structural homology search using Dali identified only 6 other structures that have a

Table 2. PCR Primers Used To Generate $PyciG_{ECO}$ for EMSA Experiments[a]

primer name	sequence (5′−3′)
$PyciG_{ECO}$-F[b]	AATTGTTAATATATCCAGAATGTTCCTCAAAATATATTTTCCCTCTAT
$PyciG_{ECO}$-R[b]	ATAGAGGGAAAATATATTTTGAGGAACATTCTGGATATATTAACAATT

[a]F indicates forward primer, and R indicates reverse primer. [b]Primer contains a 3′ biotin label.

dx.doi.org/10.1021/bi500871a | Biochemistry 2014, 53, 7223−7231

Table 3. Primers Used To Subclone *mcbR* Variants into pBS(Kan) Plasmids[a]

primer name	sequence (5′−3′)
mcbR-F	TTTGTTTGGTACCAAGAAGGAGATATACCATGGGCTCTG
mcbR-R	GCCGCAAGAGCTCATTAACGATTGTATTGCTGG

[a]F indicates forward primer and R indicates reverse primer.

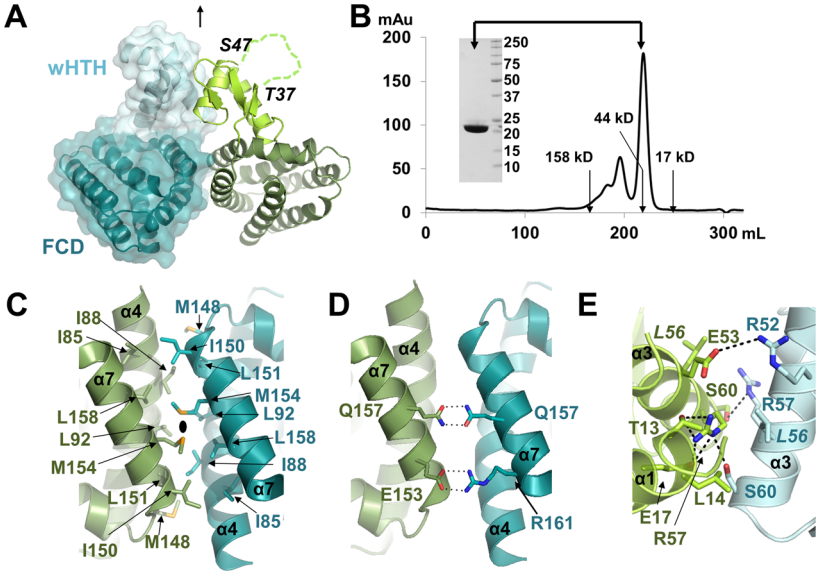


Figure 2. Dimerization interface of McbR. (A) McbR dimer, with one monomer colored in shades of blue and the second colored in shades of green. The N-terminal wHTH domains are colored in light blue and light green, and the C-terminal FCD domains are colored in teal and dark green. The residues between $\beta1/\alpha3$ of monomer B are disordered and represented as a dotted line. The pseudo 2-fold axis is indicated by an arrow. (B) Size-exclusion chromatogram of McbR with elution volumes of MW standards indicated (Bio-Rad; calculated molecular weight of the McbR monomer is ~24.5 kDa). (C) Hydrophobic interactions that stabilize the FCD domain dimerization interface; the pseudo 2-fold axis is indicated by a black circle. (D) Polar/salt bridge interactions (shown as black dashed lines) that stabilize the FCD domain dimerization interface. (E) Interactions at the wHTH domain interface (polar/salt bridge interactions shown as black dashed lines); Leu56 is labeled in italics to highlight it. Whereas $L56_A$ (light blue) is buried in the interface, $Leu56_B$ (bright green) is not and is instead at the interface periphery.

Table 4. FCD Family Members and Their Structural Similarity to McbR[a]

		FCD only			full-length						
name	PDB	*R*	*Z*	rmsd (Å)	*R*	*Z*	rmsd (Å)	Id (%)	subfamily	metal binding	FCD ligands[b]
McbR	4P9F								VanR	no	UNK
PS5454	3C7J	1	18.2	1.9	3	18.3	1.9	23	VanR	yes	Ni
CGL2915	2DI3	2	15.5	2.6	5	15.5	4.0	20	FadR	yes	Zn
RO03477	2HS5	3	15.4	2.6	2	19.7	3.1	19	VanR	no	Act
Reut_B4629	3IHU	4	15.4	2.5	1	20.3	2.6	16	VanR	no	
TM0439	3FMS	5	14.8	2.4	4	16.6	3.1	22	VanR	yes	Act, Ni
FadR	1H9G	6	12.8	2.6	6	12.6	6.2	14	FadR	no	CoA-Myr

[a]*R*, Z-score rank. *Z*, DALI Z-score. rmsd, root-mean-square deviation reported by DALI. Id, % sequence identity determined using FFAS. [b]Ligands/metals bound at the ligand binding pocket; UNK, unknown; Act, acetate ion.

high degree of similarity to McbR (*Z*-score > 7 using only the FCD domain; Table 4 and Figure 3). These represent the only other members of the FCD family with known structures. The dimerization interface mediated by the C-terminal FCD domain is topologically conserved within the FCD family and is composed of the first helix in the FCD domain ($\alpha4$ in McbR) and the N-terminal half of the kinked fourth helix ($\alpha7$ in McbR). In McbR, the FCD dimerization interface buries 1655 Å^2 of solvent-accessible surface area, which is 70% of the buried surface area (BSA) for the entire McbR dimer. The hydrophobic core of the FCD dimerization interface is formed by residues Ile85, Ile88, Leu92, Met148, Ile150, Leu151, Met154, and Leu158 from both monomers, each of which is completely occluded from solvent (Figure 2C). It is further stabilized by polar and salt bridge interactions, especially a bidentate hydrogen bond between $Gln157_A$ and $Gln157_B$ and a bidentate salt bridge between $Arg161_A$ and $Glu153_B$ (A or B subscript indicates that the residue is from subunit A or B, respectively; Figure 2D).

In McbR, the wHTH domains also interact, extending the dimerization interface beyond that typically observed in the FCD subfamily of GntR regulators. The wHTH interface buries 740 Å^2 of BSA, for a total of 2395 Å^2 buried between the two McbR monomers. Although the FCD domains are related by a near perfect 2-fold axis centered on $Gln157_{A/B}$, the wHTH are not. Instead, they are related by a rotation of ~172°. Thus,

dx.doi.org/10.1021/bi500871a | *Biochemistry* 2014, 53, 7223−7231

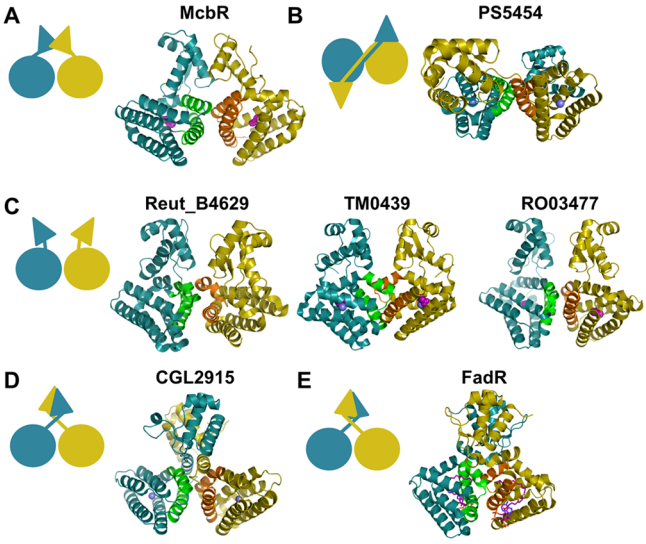


Figure 3. Quaternary structures of FCD family. FCD family members whose structures have been determined are shown, with one monomer depicted in teal and one in gold. Metals bound to the FCD domains are depicted as light blue spheres. Ligands/molecules bound in the FCD ligand binding pockets are shown as magenta spheres or sticks. The dimerization helices (α4 and α7 in McbR) are colored green and orange. The corresponding quaternary structures are depicted as cartoons, with the N-terminal domains shown as triangles and the C-terminal domains as spheres. (A) Head-to-head dimerization in which both the wHTH domains and the FCD domains contribute to the dimerization interface. (B) Dimerization in which the wHTH domains do not interact with either one another or the FCD domains. (C) Head-to-head dimerization in which only the FCD domains contribute to the dimerization. (D) Domain swapping dimerization in which the wHTH domain of one monomer reaches across the FCD domain interface to interact with the FCD domain of the second monomer. (E) Same as panel D except that the wHTH domains are inverted with respect to one another.

whereas $Leu56_A$ is buried in the wHTH interface, the corresponding residue ($Leu56_B$) is located at the interface periphery (Figure 2E). The wHTH dimerization is composed largely of polar interactions (i.e., a hydrogen bond between $Ser60_{A/B}$ and $Arg57_{A/B}$) and a few hydrophobic interactions ($Leu14_B$ and $Leu56_A$); however, unlike the residues at the FCD interface, none of the wHTH interface residues become extensively buried upon complex formation (Figure 2E). Finally, $Asn62_A$ (wHTH domain) hydrogen bonds with $Glu153_B$ (FCD domain); this is the only noncovalent interaction connecting the two different domains from the distinct subunits in the dimer.

Comparison of McbR with the other members of the FCD subfamily reveals that while the FCD dimerization interface is conserved within the family, the relative orientation of the wHTH and FCD domains is not. This gives rise to distinct differences in the orientations of the wHTH domains and, in some cases, distinct quaternary structures (Figure 3). This is why the FCD family member identified to be most similar to McbR using the DALI structural homology search database changes depending on whether the search is performed with the McbR FCD domain alone (PS5454, PDB ID: 3C7J) or full-length McbR (Reut_B4629, PDB ID: 3IHU; Table 4). As expected, the FCD proteins identified as most different from full-length McbR are FadR and CGL2915. These are also both members of the FadR subfamily, but they have an additional helix between the wHTH and the FCD ligand binding domains (Figure 4). The presence of this helix leads to domain swapped quaternary structures, in which the wHTH domain of subunit A crosses the dimerization interface to make contacts with the FCD domain of subunit B. This domain swapping is not observed in the VanR subclass of FCD regulators, and instead, in these proteins, the wHTH and FCD domains of the same subunit are more intimately associated.

The Structure of McbR Is Predominantly in a Ligand-Bound Conformation. The C-terminal FCD domains are composed of either 6 (VanR subclass) or 7 (FadR subclass) α-helices that form an antiparallel helical bundle. McbR, which has 6 helices, is a member of the VanR subclass (Figure 1A). The FCD domains have a large cavity in the center of this helical bundle, which is the location of the ligand binding site. The structure of this cavity is identical between both FCD domains in McbR, as the FCD domains superimpose with a root-mean-squared deviation (rsmd) of only 0.24 Å (Figure 6B). Although the helical topology is conserved among FCD domains, the sequence conservation among FCD family members, especially the residues that line the ligand binding cavities, is very low, likely reflecting their distinct ligand specificities (Figure 4A).

Recently, it was shown that the majority of FCD family members use three conserved histidines to bind a metal ion in the ligand binding cavity, suggesting that these regulators bind ligands that interact directly with the bound metal (Figure 4A,B).[10] In McbR, these histidines are not conserved and are instead replaced by Arg139, Tyr185, and Ile207 (Figure 4B). Thus, McbR is one of the few FCD family members that does not bind a metal. Because of this, the ligand pocket in McbR is large, with a volume of ~200 $Å^3$, nearly double that of the metal-binding FCD domains (Figure 4C). Although the endogenous ligand for McbR is still unknown, clear unambiguous density for a bound entity was observed in the FCD ligand binding cavities of both monomers of McbR (Figure 4D). None of the protein and crystallization buffer components, or derivatives thereof, fit the density. This is likely because the density is rather undefined, potentially because it is not fully occupied, a phenomenon commonly observed without externally supplied ligands and/or cofactors. Alternatively, the density could correspond to the biologically relevant ligand, as McbR is an *E. coli* protein and was expressed in *E. coli*. However, potential ligands, such as glucuronic acid, a component of colanic acid whose metabolism has been shown to be regulated by McbR, did not fit the density.[4] Finally, automated ligand fitting routines, such as the LigandFit program implemented in Phenix, also failed to identify a ligand that satisfactorily fit the density.[29,30] Because the density did not enable the identity of the ligand to be confidently determined, it has not been modeled.

However, the presence of the density did reveal the identity of the residues that likely define the McbR ligand binding site. Namely, the bound entity is strongly coordinated by two arginine residues, Arg89 and Arg139, which themselves are organized via a shared salt bridge with Glu93 (Figures 4C,D). Two neighboring asparagine residues, Asn135 and Asn211, also contribute to binding. To investigate whether these residues are important for McbR function, we generated two variants of McbR by mutating the residues that define the entity binding site. Because mutating residues in the interior of a protein can also lead to protein unfolding, we generated two distinct mutants: a double mutant in which Glu93 and Arg139 were

dx.doi.org/10.1021/bi500871a | *Biochemistry* 2014, 53, 7223−7231

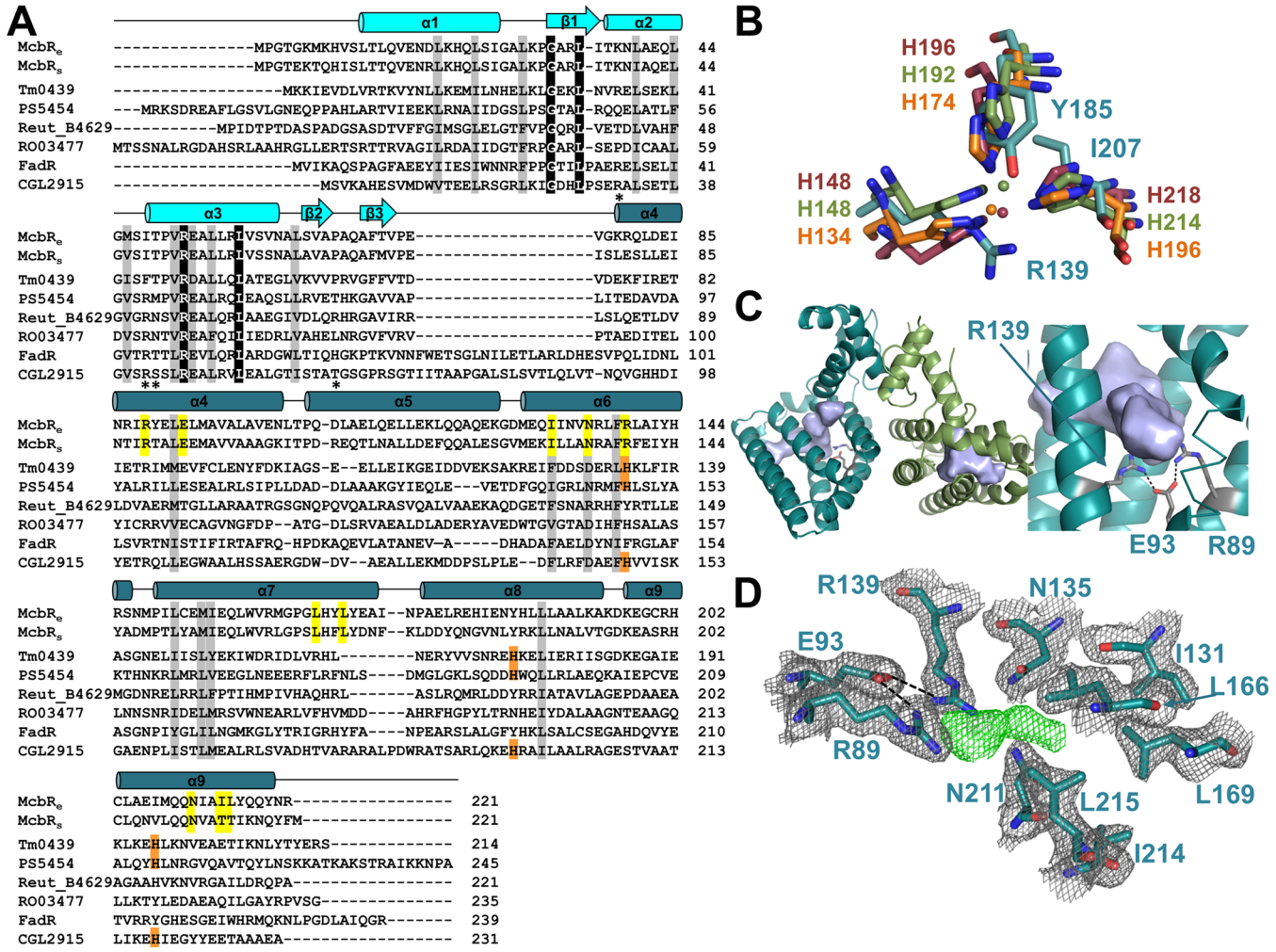


Figure 4. Ligand binding cavity of McbR and comparison to structural homologues. (A) Multiple sequence alignment showing high conservation at the N-terminal domain ($\alpha1-\beta3$) and the C-terminal domain ($\alpha4-\alpha9$) in McbR in comparison to the FCD family. Identical amino acids are highlighted in black, and similar amino acids are highlighted in gray. $McbR_e$ represents *E. coli* McbR, and $McbR_s$ represents *S. typhimurium* McbR. Alpha helices are depicted as cylinders above the sequence alignment, and the beta-strands, as arrows. Asterisks mark the residues in FadR that make base-specific contacts with DNA. Residues that define the McbR ligand binding site are highlighted in yellow. Residues important for metal binding in the FCD family are highlighted in orange. (B) Structural superposition of the three conserved histidines in metal binding FCD family members and the corresponding residues in McbR (teal). *Pseudomonas syringae* PS5454 (PDB ID: 3C7J) is shown in green, *Thermotoga maritima* TM0439 (PDB ID: 3FMS) is shown in orange, and *Corynebacterium glutamicum* CGL2915 (PDB ID: 2DI3) is shown in maroon. Each respective metal is shown as a sphere in the same color. (C) (Left) Cartoon depiction of McbR with the ligand binding cavities represented as purple surfaces. (Right) Enlarged image of the binding cavity highlighting the three conserved residues in McbR that appear to be important for ligand binding (colored as in the left panel). (D) Electron density for the ligand binding cavity in chain A of McbR. Positive density is shown as green chicken wire. Residues coordinating the unidentified entity (see text) are shown as teal sticks. Sigma level for the $2F_o - F_c$ map is 1.0. Sigma level for the $F_o - F_c$ map is 3.0.

substituted with Ser and Phe, respectively, the structurally homologous residues in FadR (the residue structurally homologous to McbR Arg89 is also an Arg in FadR) and a triple mutant in which all three residues were mutated to alanines (Arg89Ala, Glu93Ala, and Arg139Ala). CD polarimetry demonstrated that both McbR variants are folded, and EMSAs showed they are functional (Figure S1). The mutants were somewhat less thermostable (ΔT_m of −8.6 and −18.6 °C compared to WT for the double and triple mutant, respectively), but this was expected, as the mutations are in the interior of the protein; indeed, this is exactly why two mutants, one in which the residues were mutated to those present in FadR (the double mutant) and one in which the residues were simply mutated alanine (the triple mutant), were tested. McbR deletion from *E. coli* results in EPS overproduction and elicits a mucoidy phenotype.[4] This mucoidy phenotype is substantially reduced upon producing McbR ectopically (Figure 5). However, cells producing McbR with triple mutations (Arg89Ala, Glu93Ala, and Arg139Ala) are mucoid (Figure 5). This observation demonstrates the importance of Arg89, Glu93, and Arg139 in binding the unknown ligand and, more importantly, the physiological relevance of the unknown ligand in affecting EPS production. Both Arg89 and Arg139 are required for ligand binding, as cells producing McbR with only two mutations (Glu93Ser and Arg139Phe) also remain less mucoid than cells with empty plasmid or cells producing McbR with three mutations (Arg89Ala, Glu93Ala, and Arg139Ala) (Figure 5).

The Conformation of McbR Crystallized Is Likely Incompatible with DNA Binding. The wHTH domain is defined by helix $\alpha2$, a connecting turn, and helix $\alpha3$ (HTH) and a small loop in the antiparallel β-sheet (the wing). The

dx.doi.org/10.1021/bi500871a | *Biochemistry* 2014, 53, 7223−7231

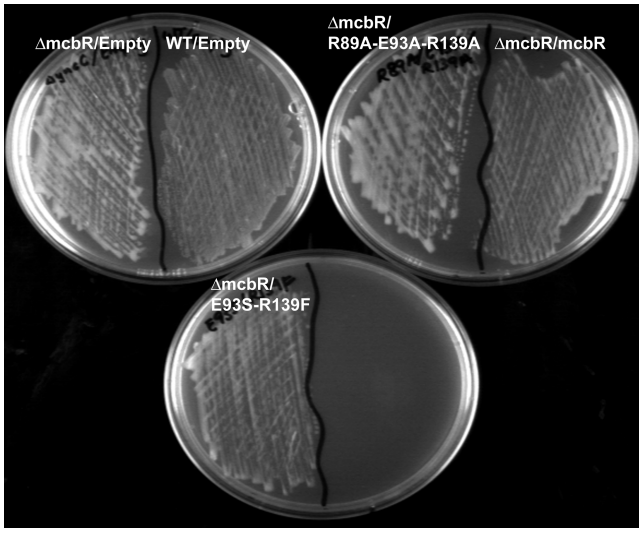


Figure 5. Mucoidy level of *E. coli* MG1655 Δ*mcbR* Δ*Km* producing different McbR variants. Each strain was grown on LB agar supplemented with 50 μg/mL kanamycin and 1 mM IPTG at 37 °C for 12 h. WT/empty, *E. coli* MG1655/pBS(Kan); Δ*mcbR*/empty, *E. coli* MG1655 Δ*mcbR* Δ*Km*/pBS(Kan); Δ*mcbR*/mcbR, *E. coli* MG1655 Δ*mcbR* Δ*Km*/pBS(Kan)-*mcbR*; Δ*mcbR*/E93S-R139F, *E. coli* MG1655 Δ*mcbR* Δ*Km*/pBS(Kan)-*mcbR*-E93S-R139F; Δ*mcbR*/R89A-E93A-R139A, *E. coli* MG1655 Δ*mcbR* Δ*Km*/pBS(Kan)-*mcbR*-R89A-E93A-R139A.

wHTH domain is slightly more conserved than the FCD domain (5% identity, 13% similar) when comparing the 7 structurally characterized FCD family members, with McbR residues Leu20, Leu24, Leu29, Gly32, Leu35, Leu40, Leu44, Met46, Val51, Arg52, Glu53, Leu55, Leu58, and Leu64 being highly similar (Figure 4A). The conserved hydrophobic residues function to stabilize the wHTH domain fold, whereas the two charged residues are located at the wHTH dimerization interface (Figure 2E).

In McbR, the C-terminal portion of wHTH helix α1 contributes to the top of the FCD binding cavity, with Ile26 (helix α1) ~12 Å away from the FCD domain ligand coordinating arginines (Arg89 and Arg139; Figure 6A). Thus, this wHTH–FCD interface provides a conduit by which effector binding in the FCD domain can be structurally communicated to wHTH DNA binding domain.[8] The conformation and orientation of the McbR wHTH domains appear to be incompatible with DNA binding. First, residues 37–48, which comprise helix α2, are disordered in subunit B (Figure 2A). Residues from helix α2 often contribute to DNA recognition, as has been observed for the FCD transcription factor FadR.[9] Second, the two domains in McbR differ not only in their relative orientations to the FCD domain but also in conformation, with an rmsd of 1.2 Å (Figure 6B). This is due to a change in the orientation of the wing between strands β2 and β3.

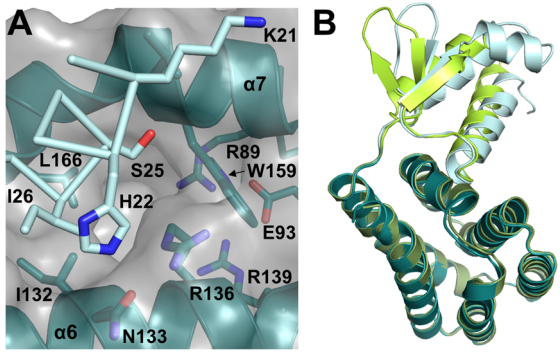


Figure 6. N-Terminal domain of McbR. (A) The pocket of α1 into the FCD domain. Helices α6 and α7 are colored teal (cartoon), and helix α1 is shown in light blue (sticks). (B) Superposition of McbR chain A (light blue/deep teal) and McbR chain B (light green/dark green). Whereas the C-terminal domains superimpose well (deep teal/dark green), the N-terminal domains (light blue/light green) do not.

Implications for McbR Function. Currently, *E. coli* FadR is the only member of the GntR family whose DNA-bound structure has been determined,[8,9] revealing that FadR binds the short palindromic consensus sequence 5′-TGGNNNNNCCA-3′. Previously, the *E. coli* McbR protein was shown to bind upstream of the *E. coli yciGFE* promoter ($PyciG_{ECO}$).[6] Subsequent DNaseI footprinting identified two distinct DNA sequences within $PyciG_{ECO}$ protected by McbR binding. To confirm that McbR binds this operator, we performed EMSA experiments using WT McbR and $PyciG_{ECO}$ DNA. As shown in Figure 7, McbR binds and shifts $PyciG_{ECO}$ DNA.

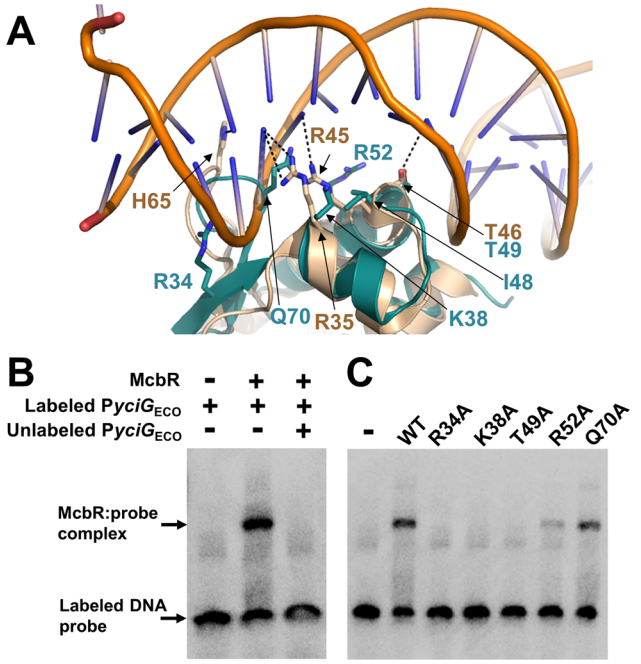


Figure 7. McbR:$PyciG_{ECO}$ EMSA experiments. (A) Superposition of the N-terminal domain of McbR (teal) and *E. coli* FadR (beige, PDB ID: 1HW2) bound to DNA. Residues making base-specific contacts in FadR and the structurally overlapping residues in McbR are shown as sticks and labeled. (B) EMSA experiments using biotin-labeled $PyciG_{ECO}$ and WT McbR (the migration of the DNA alone is shown in the left lane). (C) EMSA experiments using biotin-labeled $PyciG_{ECO}$ and either WT McbR or the McbR variants as indicated; the migration of the DNA alone is shown in the left lane. All binding reactions in panels B and C contain the nonspecific poly(dI–dC) probe.

In the FadR–DNA complex, Arg35 (helix α2), Arg45 (helix α3), Thr46 (helix α3), and His65 (β2–β3 wing) mediate base-specific contacts with the bound DNA. The corresponding residues in McbR are Lys38, Ile48, Thr49, respectively, with no residue corresponding to His65 (Figure 7A). This suggests that McbR likely interacts with DNA via helix α2 (Lys38) and helix

dx.doi.org/10.1021/bi500871a | *Biochemistry* 2014, 53, 7223–7231

α3 (Thr49). Superposition of the FadR−DNA complex and McbR shows Gln70 as the only residue with a polar side chain in close proximity with the DNA in the β2−β3 wing. Additional basic residues in close proximity to the DNA include Arg34 and Arg52 (Figure 7A). We tested the role of these residues in DNA binding using EMSA experiments performed with the P*yciG*$_{ECO}$ promoter DNA and McbR mutants (we used CD to show that the variants are folded; Figure S1; the T_m's of the variants are within 3.7 °C of that of WT, which has a T_m of 63.2 °C). The EMSA experiments show that residues Arg34, Lys38, Thr49, and Arg52 are important for DNA binding, as mutating these residues to alanine result in a loss of DNA binding compared to that of WT McbR (Figure 7B). Furthermore, Arg34, Lys38, and Thr49 have the most debilitating effects, suggesting that β1, α2, and α3 play key roles in DNA binding.

So, how is DNA binding regulated? As stated earlier, the GntR transcription factors are typically regulated by ligands that are metabolic substrates/products/cofactors of the genes that they regulate. In many cases, these genes are often located next to or near the GntR gene itself.[7] McbR was previously shown to bind the promoter of *yciGFE* and *ybiM*.[4,6] While the molecular functions of the protein products of these genes are currently unknown, *ybiM* has been shown to effect colanic acid production in a McbR-dependent manner, suggesting that colanic acid, or one of its constituents, may be the biologically relevant ligand for McbR.[4] Currently, our results suggest that this is not the case, as none of the components of colanic acid satisfactorily fit the ligand density in the McbR cavity. An examination of the genes near *mcbR* in the *E. coli* chromosome shows that they are involved in a variety of biological processes (Table 5); our data again shows that McbR is unlikely to be regulated by these metabolites (methionine, curcumin/dihydrocurcumin, iron, asparagine, and glutathione), as they also did not satisfactorily fit the density. However, sequence similarities between the *E. coli* and *Salmonella* McbR do suggest that they likely bind similar, if not identical, ligands. Namely, although the FCD domains of McbR from both organisms are less conserved than their corresponding wHTH domains (FCD domain sequence conservation: 46% identity, 74% similarity), the ligand binding residues are nearly perfectly conserved, including Arg89, Glu93, and Arg139 (Figure 4A); the only differences in the ligand binding pocket are distal from the Arg-Glu-Arg pocket: Ile214 and Leu215 (Thr214 and Thr215 in *Salmonella*). Because these residues change from hydrophobic (*E. coli*) to polar (*Salmonella*), the distal portion of the ligand may be slightly different between the organisms. Once the biologically relevant ligand(s) of McbR have been confidently identified, this ligand, or a derivative thereof, may be able to function as a novel therapeutic to target biofilms.

Table 5. DNA Sequences Surrounding the *mcbR* (*yncC*) Gene in *Escherichia coli* (MG1655)

gene	other names	gene description
yncA	*mnaT*, *b1448*	Methionine N-acyltransferase; L-amino acid N-acyltransferase
yncB	*curA*, *b1449*	Curcumin/dihydrocurcumin reductase, NADPH-dependent
yncD	*b1451*	Predicted iron outer membrane transporter
yncE	*b1452*	ATP-binding protein, periplasmic, function unknown
yncF	*ansP*, *b1453*	L-asparagine transporter
yncG	*b1454*	Glutathione S-transferase homologue
yncH	*b1455*	Conserved protein, function unknown

■ ASSOCIATED CONTENT

Supporting Information

CD spectra and EMSA of McbR variants (Figure S1). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rebecca_page@brown.edu. Phone: 401-863-6076. Fax: 401-863-9653.

Present Address

⊥(A.U.B.) Department of Chemistry, Hacettepe University, Faculty of Science, Beytepe, 06800 Ankara, Turkey.

Funding

This work was supported by a National Science Foundation Experimental Program to Stimulate Competitive Research (EPSCoR grant no. 1004057) graduate fellowship to D.M.L., an Army Research Office award to T.K.W. (W911NF-14-1-079), and a National Science Foundation CAREER Award (MCB-0952550) to R.P.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Data for this study were measured at beamline X25 of the National Synchrotron Light Source (supported principally by the Offices of Biological and Environmental Research and of Basic Energy Sciences of the United States Department of Energy and by the National Center for Research Resources of the National Institutes of Health). This research is based in part upon work conducted in the Center for Genomics and Proteomics Core Facility (with partial support from the National Institutes of Health, NCRR, grant nos. P30RR031153, P20RR018728, and S10RR02763, National Science Foundation EPSCoR grant no. 0554548, Lifespan Rhode Island Hospital, and the Division of Biology and Medicine, Brown University) and also in the Rhode Island NSF/EPSCoR Proteomics Share Resource Facility (supported in part by the National Science Foundation EPSCoR grant no. 1004057, National Institutes of Health grant nos. 1S10RR020923 and S10RR027027, a Rhode Island Science and Technology Advisory Council grant, and the Division of Biology and Medicine, Brown University).

■ ABBREVIATIONS

ASU, asymmetric unit; CD, circular dichroism; EPS, exopolysaccharide; EDTA, ethylenediaminetetraacetic acid; EMSA, electrophoretic mobility shift assay; FCD, FadR C-terminal domain; IPTG, isopropyl-β-D-thiogalactopyranoside; MPD, 2-methyl-2,4-pentanediol; PDB, Protein Data Bank; SEC, size-exclusion chromatography; SeMet, selenomethionine; TCEP, tris(2-carboxyethyl)phosphine; TEV, Tobacco Etch Virus; wHTH, winged helix-turn-helix; WT, wild-type

■ REFERENCES

(1) Lewis, K. (2001) Riddle of biofilm resistance. *Antimicrob. Agents Chemother. 45*, 999−1007.

(2) Potera, C. (1999) Forging a link between biofilms and disease. *Science 283*, 1837−1839.

(3) Spoering, A. L., and Lewis, K. (2001) Biofilms and planktonic cells of *Pseudomonas aeruginosa* have similar resistance to killing by antimicrobials. *J. Bacteriol. 183*, 6746−6751.

(4) Zhang, X. S., Garcia-Contreras, R., and Wood, T. K. (2008) *Escherichia coli* transcription factor YncC (McbR) regulates colanic acid and biofilm formation by repressing expression of periplasmic protein YbiM (McbA). *ISME J. 2*, 615−631.
(5) Whitfield, C. (2006) Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem. 75*, 39−68.
(6) Beraud, M., Kolb, A., Monteil, V., D'Alayer, J., and Norel, F. (2010) A proteomic analysis reveals differential regulation of the sigma(S)-dependent *yciGFE*(*katN*) locus by YncC and H-NS in *Salmonella* and *Escherichia coli* K-12. *Mol. Cell. Proteomics 9*, 2601−2616.
(7) Hoskisson, P. A., and Rigali, S. (2009) Chapter 1: Variation in form and function the helix-turn-helix regulators of the GntR superfamily. *Adv. Appl. Microbiol. 69*, 1−22.
(8) van Aalten, D. M., DiRusso, C. C., Knudsen, J., and Wierenga, R. K. (2000) Crystal structure of FadR, a fatty acid-responsive transcription factor with a novel acyl coenzyme A-binding fold. *EMBO J. 19*, 5167−5177.
(9) Xu, Y., Heath, R. J., Li, Z., Rock, C. O., and White, S. W. (2001) The FadR·DNA complex. Transcriptional control of fatty acid metabolism in *Escherichia coli*. *J. Biol. Chem. 276*, 17373−17379.
(10) Zheng, M., Cooper, D. R., Grossoehme, N. E., Yu, M., Hung, L. W., Cieslik, M., Derewenda, U., Lesley, S. A., Wilson, I. A., Giedroc, D. P., and Derewenda, Z. S. (2009) Structure of *Thermotoga maritima* TM0439: implications for the mechanism of bacterial GntR transcription regulators with Zn^{2+}-binding FCD domains. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 65*, 356−365.
(11) Peti, W., and Page, R. (2007) Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. *Protein Expression Purif. 51*, 1−10.
(12) Brown, B. L., Grigoriu, S., Kim, Y., Arruda, J. M., Davenport, A., Wood, T. K., Peti, W., and Page, R. (2009) Three dimensional structure of the MqsR:MqsA complex: a novel TA pair comprised of a toxin homologous to RelE and an antitoxin with unique properties. *PLoS Pathogens 5*, e1000706.
(13) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307−326.
(14) Pape, T., and Schneider, T. R. (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *Appl. Cryst. 37*, 843−844.
(15) Schreider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 58*, 1772−1779.
(16) Sheldrick, G. M. (2002) Macromolecular phasing with SHELXE. *Z. Kristallogr. 217*, 664−650.
(17) Sheldrick, G. M. (2003) *SHELXC*, Göttingen University, Göttingen, Germany.
(18) Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc. 3*, 1171−1179.
(19) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 486−501.
(20) Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 213−221.
(21) Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 12−21.
(22) Tina, K. G., Bhadra, R., and Srinivasan, N. (2007) PIC: protein interactions calculator. *Nucleic Acids Res. 35*, W473−476.
(23) Hubbard, S. J., and Thornton, J. M. (1993) *Naccess*, University of Manchester, Manchester, UK, http://www.bioinf.manchester.ac.uk/naccess/.
(24) Yu, J., Zhou, Y., Tanaka, I., and Yao, M. (2010) Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. *Bioinformatics 26*, 46−52.
(25) Canada, K. A., Iwashita, S., Shim, H., and Wood, T. K. (2002) Directed evolution of toluene *ortho*-monooxygenase for enhanced 1-naphthol synthesis and chlorinated ethene degradation. *J. Bacteriol. 184*, 344−349.
(26) Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K., and Jones, D. T. (2013) Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res. 41*, W349−357.
(27) Dosztányi, Z., Csizmók, V., Tompa, P., and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics 21*, 3433−3434.
(28) Dosztányi, Z., Csizmók, V., Tompa, P., and Simon, I. (2005) The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol. 347*, 827−839.
(29) Terwilliger, T. C., Adams, P. D., Moriarty, N. W., and Cohn, J. D. (2007) Ligand identification using electron-density map correlations. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 63*, 101−107.
(30) Terwilliger, T. C., Klei, H., Adams, P. D., Moriarty, N. W., and Cohn, J. D. (2006) Automated ligand fitting by core-fragment fitting and extension into density. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 62*, 915−922.