Published online 20 November 2017

Nucleic Acids Research, 2018, Vol. 46, No. 1 **403–420**
doi: 10.1093/nar/gkx1148

Novel structural features drive DNA binding properties of Cmr, a CRP family protein in TB complex mycobacteria

Sridevi Ranganathan[1], Jonah Cheung[2,*], Michael Cassidy[2], Christopher Ginter[2], Janice D. Pata[1,3] and Kathleen A. McDonough[1,3,*]

[1]Department of Biomedical Sciences, School of Public Health, University at Albany, SUNY, Albany, NY 12201, USA, [2]New York Structural Biology Center, New York, NY 10027, USA and [3]Wadsworth Center, New York State Department of Health, 120 New Scotland Avenue, PO Box 22002, Albany, NY 12201-2002, USA

Received July 01, 2017; Revised October 26, 2017; Editorial Decision October 30, 2017; Accepted November 13, 2017

ABSTRACT

Mycobacterium tuberculosis **(Mtb) encodes two CRP/FNR family transcription factors (TF) that contribute to virulence, Cmr (Rv1675c) and CRP_{Mt} (Rv3676). Prior studies identified distinct chromosomal binding profiles for each TF despite their recognizing overlapping DNA motifs. The present study shows that Cmr binding specificity is determined by discriminator nucleotides at motif positions 4 and 13. X-ray crystallography and targeted mutational analyses identified an arginine-rich loop that expands Cmr's DNA interactions beyond the classical helix-turn-helix contacts common to all CRP/FNR family members and facilitates binding to imperfect DNA sequences. Cmr binding to DNA results in a pronounced asymmetric bending of the DNA and its high level of cooperativity is consistent with DNA-facilitated dimerization. A unique N-terminal extension inserts between the DNA binding and dimerization domains, partially occluding the site where the canonical cAMP binding pocket is found. However, an unstructured region of this N-terminus may help modulate Cmr activity in response to cellular signals. Cmr's multiple levels of DNA interaction likely enhance its ability to integrate diverse gene regulatory signals, while its novel structural features establish Cmr as an atypical CRP/FNR family member.**

INTRODUCTION

*Mycobacterium tuberculosi*s (Mtb) is the etiologic agent of tuberculosis (TB), which causes high levels of morbidity and mortality worldwide (1). Mtb adapts to changing conditions during infection by modulating its gene expression, and the second messenger cyclic AMP (cAMP) plays a key role in Mtb's gene regulatory response to host-associated conditions (2,3). CRP_{Mt} (Rv3676) and Cmr (Rv1675c) are cAMP-associated transcription factors (TFs) in Mtb, and both are classified as CRP/FNR family TFs (4).

The CRP/FNR family of TFs is named after the two earliest discovered members of the group, *c*AMP *r*eceptor *p*rotein (CRP) and *f*umarate and *n*itrate *r*eductase (Fnr) in *Escherichia coli* (5,6). (*E. coli* CRP is referred to herein as CRP_{Ec} for clarity, and is also called CAP, for *c*atabolite *a*ctivator *p*rotein). All CRP/FNR family TFs are characterized by the presence of an N-terminal β-barrel/cyclic nucleotide binding domain, and a C-terminal helix-turn-helix (HTH) DNA binding domain (6,7). Advancement in genome analysis and phylogenetic methods has resulted in increased detection of CRP/FNR type TFs (6,8). A study classified ∼1500 CRP/FNR type regulators into 12 representative groups in 2012 (8). CRP/FNR family TFs recognize and respond to a variety of environmental signals, including cAMP, carbon monoxide, 2-oxoglutarate and cellular redox state (6,7,9–13). These TFs are also known to regulate diverse cellular processes such as catabolite repression, nitrogen fixation, metabolism, respiration, aromatic ring degradation and expression of virulence factors (6,14–18).

CRP_{Mt} is the closest Mtb ortholog of CRP_{Ec} and is better studied than Cmr (19–23). Both Cmr and CRP_{Mt} exhibit widespread genomic binding associated with the control of large putative regulons (24–26), and deletion of either gene attenuates Mtb virulence in a murine model (21,27). CRP_{Mt} regulates serine metabolism and is required for wild-type (WT) levels of growth in liquid culture and within macrophages (20,21,28). In contrast, Cmr co-regulates genes within the dormancy associated DosR regulon and contributes to nitric oxide sensitivity (26,27) but

*To whom correspondence should be addressed. Tel: +1 518 486 4253; Fax: +1 518 473 1326; Email: kathleen.mcdonough@health.ny.gov
Correspondence may also be addressed to Jonah Cheung. Tel: +1 212 939 0660 (Ext. 375); Fax: +1 212 939 0863; Email: jcheung@nysbc.org

cmr-deleted mutants are not defective for *in vitro* growth in nutrient-rich media (29).

Cmr was first identified as the regulator of a group of cAMP- and macrophage-responsive genes in Mtb (29). The genome-wide binding sites of Cmr have now been identified (26). Despite recognizing a 16-bp motif (NGTC/G-N_8-G/CACN) that fully overlaps that of CRP_{Mt}, there is minimal overlap between the genomic binding targets of the two TFs (24–26,29). Cmr binds ∼368 genomic loci in TB complex mycobacteria while CRP_{Mt} binds ∼2000 (24,26). The N-terminal region of Cmr has a putative nucleotide-sensing domain (4). cAMP affects Cmr's DNA binding enrichment at a subset of sites *in vivo*, particularly in the vicinity of genes associated with the hypoxia-inducible DosR (DevR) regulon in Mtb (26). Addition of cAMP also causes this subset of DNA binding sites to migrate anomalously when complexed with Cmr during electromobility shift assays (EMSA) (26). Nonetheless, direct binding of cAMP to Cmr has not been demonstrated and the basis for differences in DNA and cAMP binding properties between Cmr and CRP_{Mt} is not known.

CRP_{Ec} and CRP_{Mt} have been extensively studied, and serve as prototypes for understanding the CRP/FNR family of DNA binding proteins. CRP_{Mt} and CRP_{Ec} recognize and bind to the DNA motifs (C/TGTG/CA/G-N_6-T/CC/GACG/A) and (TGTGA-N_6-TCACA), respectively via their C-terminal HTH domains (20–22,24,25,30). The differences between CRP_{Ec} and CRP_{Mt} become more apparent when comparing their functional dependence on cAMP binding. CRP_{Ec} cannot bind DNA in the absence of cAMP, which causes elaborate conformational changes that position the HTH domains for DNA binding (30,31). In contrast, CRP_{Mt} shows specific DNA binding at most sites even in the absence of cAMP (19,22), although direct binding of cAMP to CRP_{Mt} enhances its DNA binding affinity ∼2-fold (20,21).

In this study, we determined a high-resolution crystal structure of Cmr, and analyzed the key features of this protein using mutational and molecular methods. Our study identifies novel structural regions in Cmr that play vital roles in determining the specificity of DNA recognition and binding. We show that accessory DNA contact regions that are not present in CRP_{Mt} influence the ability of Cmr to distinguish between sites with single nucleotide differences. Additionally, Cmr binding to DNA is cooperative and appears to bend the DNA asymmetrically. Along with its unique DNA binding features, the intriguing lack of a cAMP-binding pocket differentiates Cmr from the CRP prototypes of this family.

MATERIALS AND METHODS

Recombinant DNA construction (For X-ray crystallization)

DNA fragments encoding Cmr 1–244 and Cmr 13–244 were amplified by polymerase chain reaction (PCR) from template plasmid pMBC370, containing the *cmr* open reading frame cloned downstream of the N-terminal His6x tag in the pet28a+ vector backbone, and inserted into a modified pMCSG9 vector (32) using ligation-independent cloning (33) to create pMCSG9 Cmr1 and pMCSG9 Cmr13. This fused a decahistidine-tagged maltose-binding-protein (His_{10}MBP) at the N-terminal of Cmr that could be cleaved using tobacco etch virus (TEV) protease.

Protein expression and purification (For X-ray crystallization)

Expression vectors pMCSG9 Cmr1 and pMCSG9 Cmr13 were transformed into *E. coli* BL21(DE3) cells. Using the autoinduction method (34), initial cultures in 100 mM phosphate non-inducing minimal media (PAG) containing 100 μg/ml carbenicillin were grown overnight at 37°C and then subsequently inoculated into autoinducing PA-5052 or PASM-5052 media (100 μg/ml carbenicillin added in both) for expression of either unlabeled native protein or selenomethionyl(SeMet)-protein for structure determination, respectively. Fusion proteins were expressed overnight at 30°C and purified via two-step nickel-affinity/gel filtration chromatography using a previously described method (35).

TEV protease (36) was used to cleave the MBP-Cmr fusion proteins at room temperature. The NaCl concentration was subsequently raised to 300 mM and cleaved MBP and TEV were removed using a 1 ml HiTrap SP column (GE Healthcare), which flowed through without binding. The ion exchange wash buffer consisted of 20 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES) pH 7.5, 200 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA) and 0.3 mM tris(2-carboxyethyl)phosphine (TCEP), and the ion exchange elution buffer was the same except the concentration of NaCl was 1 M. Purified Cmr 1–244 and Cmr 13–244 was eluted using either a NaCl gradient or a 650 mM NaCl step. Purified unlabeled protein for biochemistry was dialyzed into buffer containing 20 mM 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris) pH 8.0, 150 mM NaCl, and 0.3 mM TCEP. SeMet-labeled protein was exchanged into crystallization buffer (10 mM HEPES pH 7.5, 150 mM NaCl, 0.1 mM EDTA and 0.3 mM TCEP) using gel filtration chromatography (GE Healthcare HiLoad 16/60 Superdex column) and concentrated to 6–8 mg/ml prior to crystallization.

Crystallization and data collection

Crystals from which structures were determined did not require optimization and were grown using the sitting-drop vapor diffusion method directly from initial sparse matrix screening in which 0.2 μL of protein was mixed with an equal volume of crystallization buffer from commercial JCSG Core Suite I, II, III, and IV screens (Qiagen). At room temperature, Cmr 1–244 crystallized in a condition containing 20% (wt/vol) polyethylene glycol (PEG) 3350 and 0.2 M potassium sulphate, and Cmr 13–244 crystallized in a condition containing 1 M di-ammonium hydrogen phosphate, 0.1 M imidazole pH 8.0 and 0.2 mM NaCl. Crystals were flash frozen in liquid nitrogen after briefly soaking in a cryoprotectant solution containing the crystallization buffer supplemented with 18–20% (vol/vol) ethylene glycol. A dataset from a single crystal of each Cmr 1–244 and Cmr 13–244 was collected at the X4C beamline, tuned to the Se K-edge, at the National Synchrotron Light Source of Brookhaven National Laboratory. For Cmr 13–244, both a

high-angle and low-angle dataset was collected and subsequently combined.

Structure determination and refinement

X-ray diffraction data was indexed and merged using HKL2000 (6) and CCP4 (37). Phases for the structure of Cmr 13–244 were calculated by single-wavelength anomalous dispersion (SAD) (38) using the AutoSol script in Phenix (39) and yielded interpretable density-modified electron density maps in space group $P2_12_12$. ARP/wARP 7.5 (40) of the CCP4 package was used to build the initial model and COOT (41) and Phenix were used for final manual model building and refinement. The structure of Cmr 1–244 was solved using Cmr 13–244 as a molecular replacement search model using Phaser (42). Refinement of Cmr 1–244 was performed in a similar manner as Cmr 13–244. For both structures, a random 5% of reflections were used as the cross-validation set. Model validation statistics were calculated using Molprobity (43).

Motif analysis

The 30-nt sequences flanking the center of Cmr or CRP_{Mt} binding sites (total length of sequences = 60 nt) were extracted from the H37Rv genome and used as input sequences for MEME motif analysis (44) (http://meme-suite.org/tools/meme). The 359 Cmr binding sites, and 1000 CRP_{Mt} binding sites were considered for this analysis. The program was not restricted to search for palindromic motifs.

Cloning, expression and purification of proteins (for *in vitro* assays)

Recombinant His-tagged *cmr* (full length) and the HTH+ and HTH− truncated *cmr* were PCR amplified from *M. tuberculosis* H37Rv DNA template, using primers listed in Supplementary Table. S1. The amplified DNA products were cloned into pET28a+ (Novagen) between EcoRI and HindIII restriction sites to generate pMBC370, pMBC787 and pMBC788 respectively, all with a N-terminal His6x tag. C-terminal His6x tagged Cmr (pMBC1761) was expressed by cloning *cmr* ORF (open reading frame) into the NcoI and XhoI sites in the pET28a+ vector. Mutant proteins CmrR28A, CmrR35AR39A, Cmr-Arg_{loop}* and CmrNΔ12 were constructed using sequence overlap extension primers listed in Supplementary Table. S1, and cloned into the NcoI and XhoI sites in pet28a+ vector to create strains pMBC1781, pMBC1782, pMBC1783 and pMBC1916 respectively. The plasmids were sequence verified, and maintained in *E. coli* BL21 (DE3) strain.

Bacterial cultures were grown to OD_{600nm} 0.4–0.6, and the expression of full length Cmr was induced with 1 mM isopropyl-β-thiogalactopyranoside (IPTG) for 3 h at 21°C. The expression of the truncated proteins was induced overnight at 15 °C with 1 mM IPTG. Protein expression was confirmed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and western blot analysis using anti-His monoclonal antibody (Clontech).

Protein purification was carried out as described previously (26). For purification, a 500-ml culture was pelleted, washed with 1 mM Tris–HCl, and resuspended in 12.5 ml lysis buffer containing 50 mM Tris–HCl (pH 8.0), 0.02% glycerol, 1 mM dithiothreitol and 1% protease inhibitor cocktail (Sigma). Bacteria were lysed by three freeze-thaw cycles and sonication for 5 min (at 4°C), followed by three additional freeze-thaw cycles. The lysate was cleared by centrifugation at 13 000 rpm for 15 min at 4°C. The His-tagged proteins were purified using HisTrap affinity column (GE Lifesciences) per the manufacturer's instructions, and the eluted protein was dialyzed against phosphate-buffered saline with 10% glycerol. Protein concentration was measured with a NanoOrange Protein Quantitation Kit (Molecular Probes) and diluted to 0.3 mg/ml before being stored in aliquots at −70°C.

Cross-linking of Cmr by glutaraldehyde

Cross-linking of purified His-tagged Cmr was performed as previously described (19). His-Cmr was diluted to 310 nM in cross-linking buffer (50 mM sodium phosphate, pH 7.4, 20% glycerol, 5 mM $MgCl_2$), and was then incubated with glutaraldehyde at a final concentration of 7 mM for 1 h at room temperature. The reaction was quenched by the addition of SDS-PAGE sample buffer, and 8 μl of cross-linking reaction was separated on a 12% SDS-PAGE gel. The protein was transferred onto a PVDF membrane and visualized by Western blot with anti-Cmr serum.

Electrophoretic mobility shift assays (EMSA)

PCR forward primers were labeled with [γ- ^{33}P]-ATP (MP Biomedicals or Perkin Elmer) using T4 DNA polynucleotide kinase (New England Biolabs). DNA probes were generated using labeled forward primer and unlabeled reverse primer in PCR reaction. About 0.05 pmol DNA probe was used in each 10 μl binding reaction mixture. Briefly, 0.3 μM His6x-Cmr (C-terminal tagged) and DNA probes were incubated at room temperature for 30 min in DNA binding buffer [10 mM Tris-HCl (pH 8.0), 50 mM KCl, 1 mM EDTA, 50 μg/ml bovine serum albumin, 1 mM dithiothreitol, 0.05% non-ionic P-40 detergent, 20 μg/ml poly(dI-dC) and 10% glycerol], with or without 100μM cAMP. Samples were loaded on a 6, 8 or 12 non-denaturing polyacrylamide gel, depending on the size of the DNA probe, and run for 2–3 h at 14 V/cm in 0.5× Tris-borate-EDTA buffer at 4°C. Gels were transferred to Whatman paper, vacuum-dried, exposed overnight on a phosphor screen, scanned with Storm 860 PhosporImager (Molecular Dynamics), and analyzed with ImageQuant software (Molecular Dynamics).

To estimate the DNA binding affinity (K_d) of the proteins, EMSA experiments were performed by titrating increasing amounts of purified proteins into the DNA binding reaction, while keeping the concentration of radiolabeled DNA constant. The fraction bound in each lane was quantified using densitometry (ImageQuant, Molecular Dynamics). K_d was calculated using nonlinear-regression analysis and curve fitting using 'Specific binding- One site with Hill slope' model in GraphPad Prism software. K_d and Hill coefficient (h) were calculated using the formula

$$Y = \frac{B_{max} * X^h}{K_d^h + X^h}$$

where, 'Y' denotes specific fraction bound, 'B_{max}' denotes maximum specific binding, 'X' denotes ligand concentration, 'K_d' denotes binding affinity and 'h' denotes Hill coefficient.

RESULTS

Comparison of genome-wide binding profiles of Cmr and CRP_{Mt}

Genome-wide binding sites for Cmr and CRP_{Mt} were recently identified in TB complex mycobacteria using ChIP-seq (24–26). Their genomic binding profiles were distinct, with CRP_{Mt} binding nearly five times as many sites as Cmr bound (2000 versus 368, respectively) (24,26). However, MEME motif analysis (44) of the binding sites showed that the core 16-bp DNA binding motif recognized by Cmr (NGTC/G-N_8-G/CACN) includes that of CRP_{Mt} (C/TGTG/CA/G-N_6-T/CC/GACG/A) (Supplementary Figure S1a) (20,24,26). Mutational analyses showed complete loss of *in vitro* DNA binding by either TF upon individually mutating any of the four critical nucleotide positions (2G, 3T, 14A and 15C), common to their binding sites (22,26). Nucleotide positions 4G and 13C were previously shown to also be critical for CRP_{Mt} binding *in vitro* (20,23). Therefore, it was surprising that Cmr bound fewer genomic sites than CRP_{Mt} despite binding a DNA motif that appears to be less specific than the CRP_{Mt} binding motif.

We compared the genomic DNA binding profiles of Cmr and CRP_{Mt} by searching within published ChIP-seq datasets for the presence of CRP_{Mt} binding sites within 500 bp of the center of each Cmr binding site (24,26). A total of 189 out of the 359 (52.6%) Cmr binding sites considered in this analysis had at least one CRP_{Mt} binding site within 500 bp (Figure 1A), and 95 (26.4%) Cmr binding sites were within 50 bp of a CRP_{Mt} binding site (Figure 1B). While some of these TF binding sites directly overlap, the majority do not. A total of 266 of 1700 (15.7%) CRP_{Mt} binding sites considered in this analysis were within 500 bp, while 107 (6.3%) mapped within 50-bp of a Cmr binding site (Figure 1C). Thus, despite recognizing seemingly identical DNA binding motifs, Cmr and CRP_{Mt} bind very distinct sites in the Mtb genome. While these data raise the possibility of regulatory interactions between Cmr and CRP_{Mt} at some sites, they strongly suggest that Cmr has additional determinants of specificity that are not obvious from the binding motif.

Motif positions 4 and 13 determine Cmr versus CRP_{Mt} binding specificity

We extended the motif analysis to include sequences flanking the 16-bp core binding sites, but did not identify any additional highly conserved nucleotide positions (Figure 2A). Therefore, we focused on the core DNA binding motifs of the two TFs to understand the molecular basis of Cmr versus CRP_{Mt} motif recognition specificity. We tested the ability of Cmr to bind a well-characterized CRP_{Mt} binding site in the intergenic region of *serC*-Rv0885 (28) that contains all known critical nucleotide positions required for Cmr binding (Supplementary Figure S1a and Figure 2B). Despite the presence of all the conserved nucleotides at positions 2, 3, 14 and 15, Cmr bound this DNA sequence very weakly (Figure 2C). Single point mutations at nucleotide positions 4(G: C) or 13(C: G) in the CRP_{Mt} binding site resulted in robust Cmr binding to these DNA sequences, and complete loss of CRP_{Mt} binding. Double mutation to 4(G: C) and 13(C: G) further improved Cmr binding to the DNA sequence. Conversely, Cmr binding was decreased by altering nucleotides at positions 4 or 13 in the PRv1675c Cmr binding site to make it better match the conserved CRP_{Mt} binding motif (Figure 2D). Double mutation of this DNA fragment at motif positions 4 and 13 resulted in complete loss of Cmr binding.

DNA sequences flanking a core binding site can affect the ability of TFs to recognize and bind DNA, despite the lack of conserved sequence motifs in our extended MEME analyses (45–47). To assess the effect of flanking sequences on the observed differences in DNA binding, we tested the ability of Cmr to bind a hybrid Cmr ($flank_{CRP}$) DNA fragment, which contains the WT 16-bp core palindromic Cmr binding site from PRv1675c and flanking sequences from the CRP_{Mt} binding Rv0885c-*serC* DNA region. Similarly, we the tested the ability of CRP_{Mt} to bind a hybrid CRP_{Mt} ($flank_{Cmr}$) DNA fragment containing the WT 16-bp palindromic CRP_{Mt} binding site from Rv0885-serC CRP_{Mt} binding region, with flanking sequences from PRv1675c. Both Cmr and CRP_{Mt} bound the DNA sequences containing their respective core binding nucleotides regardless of the substituted flanking sequences (Figure 2C and D). Together, these results indicate that nucleotide positions 2, 3, 14 and 15 comprise a core DNA binding sequence needed for binding by both TFs, while positions 4 and 13 determine the specificity of TF binding. Flanking sequences do not affect binding specificity when both discriminator positions contain Cmr's preferred nucleotides (4C and 13G).

Determination of Cmr structure

We solved the crystal structure of Cmr to elucidate factors that differentiate Cmr and CRP_{Mt} DNA binding at the atomic level. *S*ingle-wavelength *a*nomalous *d*ispersion (SAD) was used to solve an initial structure of Cmr 13–244, which was then used for molecular replacement to solve the structure of full length Cmr 1–244 at 1.85Å resolution (Figure 3A). Data collection and refinement statistics are listed in Table 1. In Cmr 1–244, two molecules (A and B) formed a dimer in the asymmetric unit. The structure of Cmr 13–244 showed a similar dimer with the dimer axis coincident with the crystal axis. In both structures, residues at the N-terminal were disordered. Cmr 13–244 was ordered from position 21 to 244, while Cmr 1–244 was ordered from positions 20 and 19 to 244, in molecules A and B respectively. In Cmr 1–244, two ordered sulfate ions present in the crystallization buffer form lattice contacts (not shown). Only small differences were observed upon superposition of molecules A and B of Cmr 1–244 with each other (all C_α RMSD = 0.655Å), and with Cmr 13–244 (all C_α RMSD = 0.772Å and 1.067Å with molecules A and B, respectively). These differences may be related to crystal packing. The structure of Cmr 1–244 was used for further detailed analysis, as it was more complete.

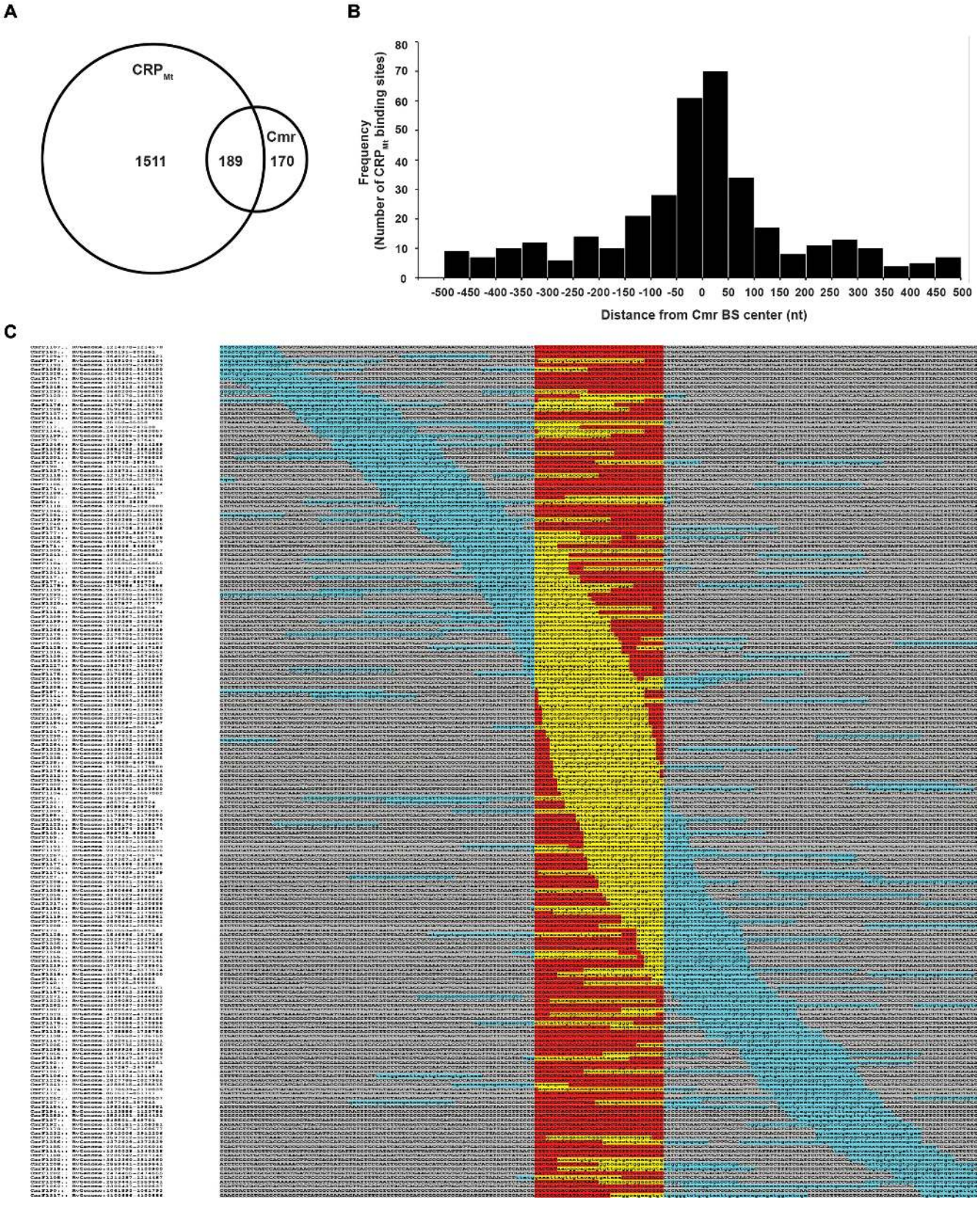

Figure 1. Comparison of genome-wide binding profiles of Cmr and CRP_{Mt}. (**A**) Venn-diagram showing the number of Cmr binding sites containing at least one CRP_{Mt} binding site within 500-bp of its center. A total of 359 Cmr binding sites were analyzed for the presence of such proximal CRP sites. (**B**) Histogram showing distribution of CRP_{Mt} binding sites within 500-bp sequences of a Cmr site (calculated center to center; and (**C**) Cmr binding sites that contain at least one CRP_{Mt} binding site within 100-bp from their center are shown. Cmr binding sites are in lowercase and highlighted in red; CRP_{Mt} binding sites are highlighted in cyan; overlapping binding site sequences are highlighted yellow.

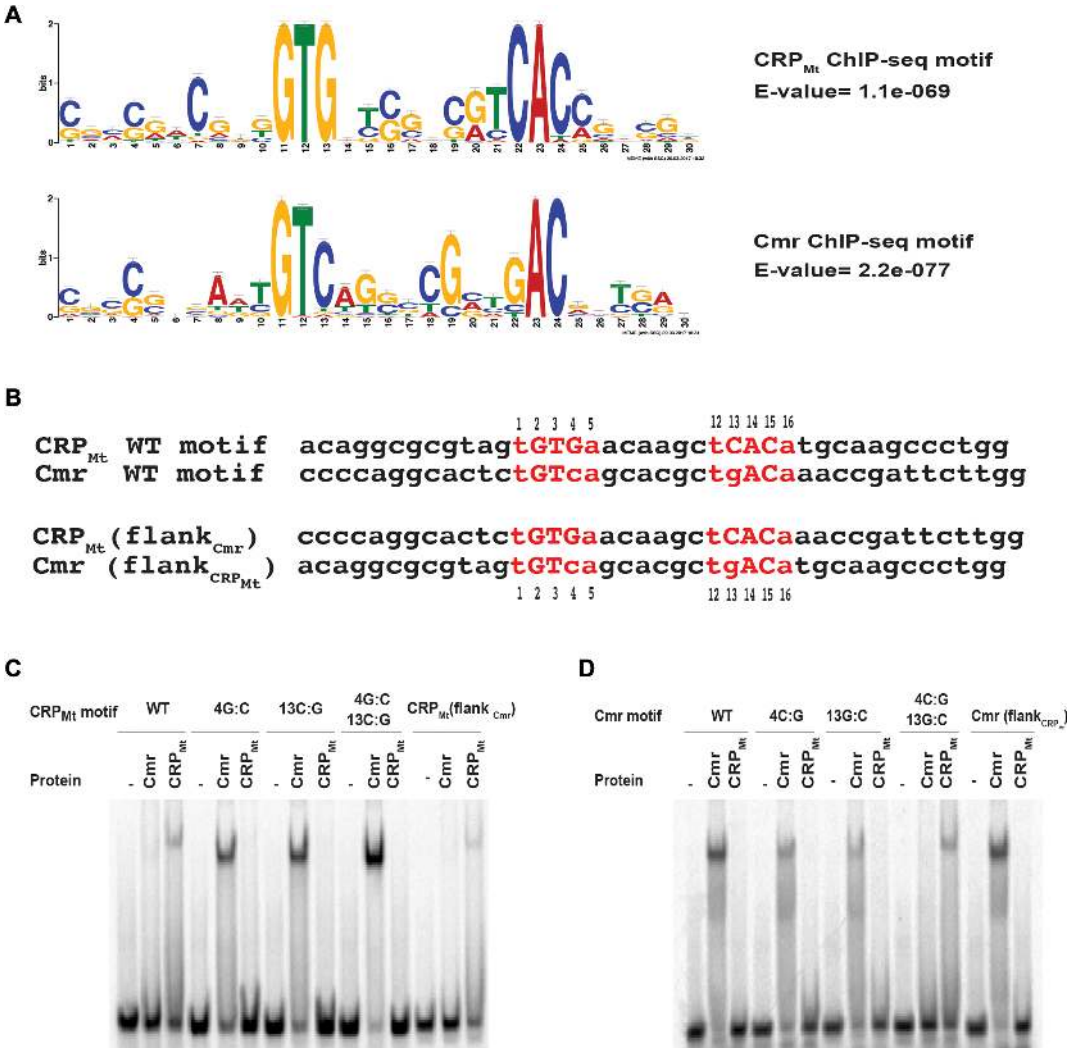

Figure 2. Motif specificity of Cmr and CRP_{Mt}. (**A**) Extended MEME motif analyses using 30-bp flanking sequences centered at the CRP_{Mt} (24) and Cmr (26) binding sites. See 'Materials and Methods' section for details on parameters used. (**B**) 40-mer sequence containing Cmr binding site from PRv1675c and CRP_{Mt} binding site from Rv0885-serC intergenic region. The palindromic half sites are shown in red. The sequences of flank-substituted DNA fragments are also shown. EMSA showing the DNA binding ability of CRP_{Mt} and Cmr to WT and mutated (**C**) CRP_{Mt} binding site and (**D**) Cmr binding site.

A comparison of Cmr with CRP/FNR family transcription factors

A search using Dali (48) identified members of the CRP/FNR transcriptional regulator family such as *Corynebacterium glutamicum* GlxR (PDB ID 4BYY) and Mtb CRP_{Mt} (PDB ID: 3H3U) as structural homologs of Cmr. GlxR and CRP_{Mt} share relatively low amino acid sequence identity with Cmr (26 and 24%, respectively) (4,49,50), while GlxR and CRP_{Mt} share 79% sequence identity with each other. Sequence alignments were generated in ClustalW and related figures were colored using ESPript 3.0 (51) (Figure 3B). Cmr only shares ∼15% amino acid level sequence identity with *E. coli* Fnr protein, another prototype member of the CRP-FNR superfamily. Furthermore, Cmr has an extended N-terminal region that is not conserved (Figure 3B).

In this study, we used the structure of CRP_{Mt} in complex with cAMP and DNA (PDB ID: 3MZH) for all structural comparisons with Cmr 1–244, and our naming of secondary structure elements in Cmr 1–244 (Figure 3A and B) follows that of CRP_{Mt} (50). Cmr and CRP_{Mt} can be aligned by superimposing 140 corresponding C_α positions to yield a RMSD of 1.92Å (Figure 3C). Similar to CRP_{Mt}, each subunit of Cmr contains a large N-terminal domain comprised of a β-sheet core, ending with helix B and joined by central helix C to the smaller C-terminal DNA-binding domain which contains the characteristic HTH domain with DNA-binding helices E and F (Figure 3A and C).

Differences between Cmr and CRP_{Mt} structures

Despite similarities in overall fold, there are significant differences between the structures of Cmr and CRP_{Mt}. The C-terminal domains of Cmr and CRP_{Mt} show the most conservation (Figure 3B) and superimpose with RMSD value of 1.18Å over 67 corresponding C_α positions (Figure 3A). The N-terminal domains of Cmr and CRP_{Mt} are less con-

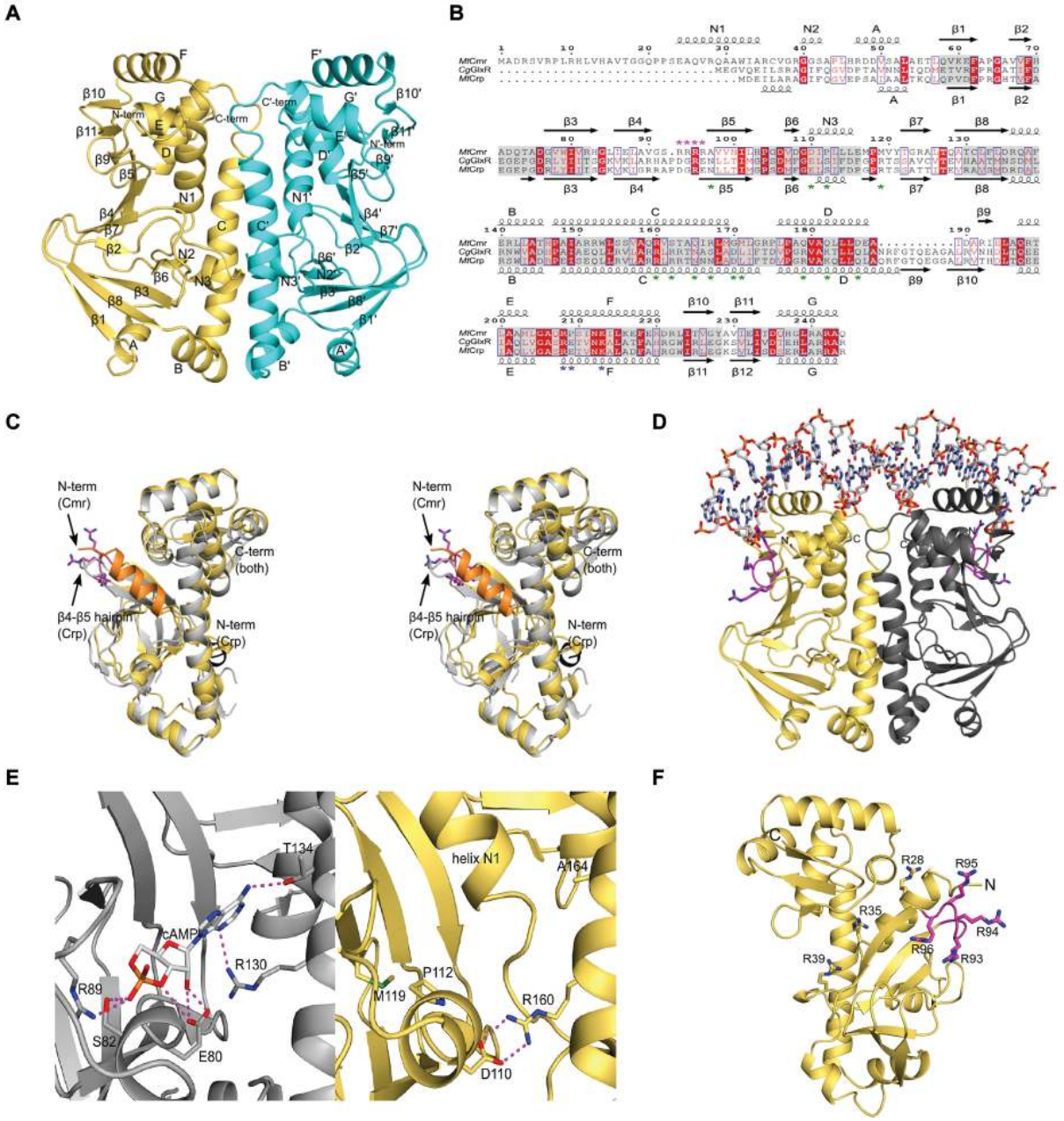

Figure 3. Structure-based alignment and structural superposition of Cmr and CRP_{Mt}. (**A**) The structure of Cmr 1–244 is shown as a ribbon diagram with secondary structure elements labeled. As residues from the N-terminal end were disordered, subunit A (colored yellow) shows amino acid residues from Q20 to the C-terminal end and subunit B (colored turquoise) shows residues from G19 to the C-terminal end. Helices E, F, E' and F' constitute the two HTH DNA binding motifs of Cmr; (**B**) The sequences of *Mycobacterium tuberculosis* Cmr, *Corynebacterium glutamicum* GlxR and *M. tuberculosis* CRP_{Mt} are aligned (adapted from ClustalW and colored using ESPript 3.0 (51). Labeled secondary structure elements for Cmr and CRP_{Mt} are shown above and below the sequences, respectively. Conserved residues are shown in red; gray boxes define residues in alignment, where at least three contiguous Cα positions are within 3.5Å of each other. The green asterisks mark the amino acid residues in CRP_{Mt} that directly contact cAMP and form the binding pocket. The blue asterisks mark residues of CRP_{Mt} involved in DNA binding specificity. The magenta asterisks mark residues of the Arg-loop in Cmr; (**C**) In stereo, subunits of Cmr 1–244 (yellow) and CRP_{Mt} (light gray) are drawn as ribbon diagrams and are superimposed. The Cmr subunit only shows amino acid residues from Q20 to the C-terminal end. The first helix of Cmr and CRP_{Mt} are colored orange and black, respectively. The carbon backbone of the Cmr Arg-loop (R93-R96) is colored magenta, arginine side chains drawn as sticks are colored magenta and blue for carbons and nitrogens, respectively; (**D**) A model of a Cmr 1–244 dimer bound to DNA, based on PDB structure: 3MZH, is shown with protein and nucleic acid atoms drawn as ribbons and sticks, respectively. The subunits of Cmr are colored yellow and dark gray, and the arginine loops are colored in the same scheme as in Figure 3C; (**E**) A close up view of the cAMP-binding pocket of CRP_{Mt} (gray ribbons and carbons) showing side chain coordination to cAMP is shown on the left. The corresponding region of Cmr (yellow ribbons and carbons) is shown on the right from the same vantage point. Side chains and cAMP are drawn as sticks. Nitrogen, oxygen, phosphorus and sulfur atoms are colored blue, red, orange and green respectively; (**F**) The structure of Cmr1–244 is shown as a yellow ribbon and mutated arginines are drawn as sticks. Nitrogens are colored blue and carbons are colored yellow except in the Arg-loop (R93-R96) where they are colored magenta.

Table 1. Data collection and refinement statistics

Data collection		
	Cmr 1–244	Cmr 13–244
Beamline	NSLS X4C	NSLS X4C
d_{min} (Å)	1.85	1.80
wavelength (Å)	0.9792	0.9792
No. of observations	190166	107416
multiplicity[a]	4.5 (3.6)	5.2 (3.1)
average $(I)/(\sigma_I)$[a]	17.6 (2.3)	19.6 (1.4)
completeness[a] (%)	99.9 (97.8)	99.6 (94.4)
R_{merge}[a,b] (%)	7.0 (53.5)	5.7 (66.0)
$CC_{1/2}$[a]	0.997 (0.776)	0.999 (0.738)
CC^{*}[a, c]	0.999 (0.931)	1.000 (0.918)
Refinement		
Bragg spacings (Å)	41.3–1.85	43.9–1.8
space group	$P2_12_12$	$P2_12_12$
cell parameters: *a,b,c* (Å)	51.62, 92.24, 102.76	55.11, 72.53, 53.94
R^{d}/R_{free}[e] (%)	17.4 (21.2)	17.3 (21.8)
No. of unique reflections	80 614	38 639
No. of total atoms (non-H)	3901	1828
No. of protein atoms (non-H)	3552	1737
No. of hetero atoms (non-H)	11	1
No. of waters	338	90
average B-factor (Å^2)	34.5	45.6
rmsd bond length (Å)	0.011	0.010
rmsd bond angle (°)	1.27	1.29
Ramachandran favored/allowed[f] (%)	99.3/0.7	100.0/0.0
Z_a	2	1
PDB code	5W5A	5W5B

[a]Values in outermost shell are given in parentheses.
[b]$R_{merge} = (\sum I_i - <I_i>)/\sum I_i$, where I_i is the integrated intensity of a given reflection.
[c]$CC^{*} = (2CC_{1/2}/1+ CC_{1/2})^{1/2}$ [50]
[d]$R = \sum|F_o| - |F_c|/\sum|F_o|$, where F_o and F_c denote observe and calculated structure factors, respectively.
[e]R_{free} was calculated using 5% of data excluded from refinement.
[f]Calculated using Molprobity (43).

served than the C-terminal domains, and superimpose with RMSD value of 1.42Å over 88 corresponding C_α positions. The extended N-terminal region of Cmr contains helix N1, which is not conserved and adopts a very different topology compared to CRP_{Mt} (Figure 3A and C). It should be noted that 19 amino acid residues at the N-terminus of Cmr were disordered and are not shown in Figure 3A, C, D and F. Residues 1–19 in Cmr 1–244, which are disordered in the structure, are also predicted to be disordered by RONN (52) and DISOPRED (53). Helices N1 and N1′ of Cmr 1–244 lie between the core β-sheet scaffold of the N-terminal domain and corresponding helices C and C′, respectively. The position of helices N1 and N1′ implies that the disordered residues at the N-terminus would be located at or near the surface of the protein where DNA binding occurs, although their function is not known. Helices N1 and N1′ also form part of the dimer interface between the two subunits wherein self-association mediated by contacts between helices N1, N3, B, C of subunit A with N1′, N3′, B′ and C′ of subunit B buries 1388Å^2 of total surface area. In contrast, the first helix of CRP_{Mt} is not involved in dimerization. When Cmr and CRP_{Mt} are superimposed, helix N1 of Cmr occupies a location corresponding with the β4–β5 hairpin of CRP_{Mt} (Figure 3C).

In each subunit of CRP_{Mt}, cAMP is found bound in a pocket formed between helix C and the N-terminal domain β4–β5 hairpin loop (54). A similar pocket is notably absent in Cmr 1–244 (Figure 3E). Instead this region of Cmr is transversed by helix N1 which forms core interactions with helix C and the β4–β5 hairpin. Additionally, side chains (S82, R89 and T134) that allow for hydrogen bonding and coordination of cAMP in CRP_{Mt} are not conserved in Cmr where instead the corresponding residues (P112, M119 and A164) are hydrophobic. A clearly identifiable ligand-binding pocket is not observed elsewhere in Cmr 1–244.

A solvent exposed 'arginine loop' (hereafter referred to as Arg-loop) located between β-strands 4 and 5 of Cmr contains four consecutive arginine residues at positions 93–96. The sequence of this loop is not conserved in CRP_{Mt}, wherein the sequence is DGRE (Figure 3B), and the location of this loop is also not conserved due to different relative positions of β4 and β5 in the two proteins. Another noteworthy detail is that in Cmr 1–244, the sequence between helices D and E is ten residues shorter than in CRP_{Mt}, forming just one strand where there are two in CRP_{Mt}. The functional significance of this difference is not known.

Modeling DNA binding to Cmr 1–244

A model of Cmr bound to DNA (Figure 3D) was constructed by superimposing a subunit of Cmr 1–244 onto each subunit of CRP_{Mt} in the DNA-bound complex (PDB ID: 3MZH) without significant disruption of the Cmr 1–244 dimeric structure. In this model, the position of one of

the phosphates in the DNA backbone coincides with a sulfate ion bound between helix E and the β10–β11 loop in Cmr 1–244 (not shown). Arginine residues are known to interact with the DNA backbone and the minor groove (55), and this model places the Arg-loops of each Cmr subunit in proximity to regions of the DNA flanking the recognition sequence, suggesting that they may be involved with DNA binding. Arg loop-DNA interactions beyond those formed with the HTH region might explain the larger DNase I footprint of Cmr (34-bp) (26) compared to that of CRP_{Mt} (27-bp) (19) (Supplementary Figure S1b). Other HTH TFs have been shown to interact with DNA through regions of the protein in addition to the HTH motif (56). Disordered residues 1–19 may become ordered upon binding to a partner and serve a role in regulation (57) as they are also located in a position that could allow DNA interactions (Figure 3D).

Additional regions of Cmr are involved in DNA binding

Based on their predicted locations proximal to the DNA binding face of Cmr (Figure 3D), we reasoned that the N-terminal region and the Arg-loop might contribute to DNA binding. We also considered the effects of residues R28, R35 and R39 in helices N1 and N2, as this region shows the most variation from the CRP_{Mt} structure, and is analogous to the cAMP-binding pocket in CRP_{Mt}. Site-specific mutations were made targeting each of these regions to determine their contributions to DNA binding (Figure 3F).

C-terminal His6x-tagged WT Cmr and mutant proteins-CmrR93–96A (denoted hereafter as Cmr-Arg_{loop}*), CmrR28A, CmrR35AR39A and CmrΔN12 were overexpressed and purified from *E. coli*. The majority of Cmr WT and mutant proteins appeared as ∼30 kDa bands in a western blot assay, consistent with the predicted molecular weight of the monomeric form of the proteins (Figure 4A). Another band migrating at ∼60 kDa was also observed with the WT and mutant proteins, and likely represents a dimer. Chemical crosslinking with glutaraldehyde increased the proportion of dimeric and higher molecular weight forms of the protein, indicating that all of the mutant proteins retained the ability to homodimerize in solution and that functional differences observed between the proteins due to large scale destabilization of protein structure are unlikely. However, we note that, unlike the N-terminal region and the Arg-loop, residues R28, R35 and R39 are involved in key interactions between structural elements in the protein. Thus, any mutations to these residues may alter DNA binding through more subtle perturbations of structure in a way that still preserves dimerization. Therefore, we have exercised cautious restraint in inferring results from the R28A and R35A39A mutant proteins, and focused primarily on the N-terminal and Arg-loop regions in this study.

Arg-loop is required to bind 'imperfect' motifs

We tested the ability of the recombinant proteins to bind a 40-bp dsDNA fragment containing a Cmr binding site from the promoter region of *cmr* (PRv1675c; same DNA fragment used in Figure 2D). Along with the critical core nucleotides (2G, 3T, 14A and 15C), this binding site also contains nucleotides 4C and 13G that are preferred by Cmr for robust binding (Figure 2D). Cmr WT and all mutant proteins bound the PRv1675c DNA probe (Figure 4B). Next, we compared the binding of the recombinant proteins to a 40-bp dsDNA fragment containing the Cmr binding site from the Rv3063 intragenic region. This binding site contains all the critical core nucleotides but has a less preferred 13C base, which we expected to negatively affect Cmr's DNA binding ability (Figure 2D). Robust binding was observed with Cmr WT, CmrNΔ12 and CmrR28A proteins, but Cmr R35AR39A displayed very weak binding and the Cmr-Arg_{loop}* protein failed to bind this DNA fragment (Figure 4B). None of the proteins bound the negative control fragment in which the Cmr binding site was scrambled.

Based on the ability of Cmr-Arg_{loop}* to bind PRv1675c but not the Rv3063 DNA fragment, we reasoned that the Arg-loop might facilitate Cmr binding to DNA sequences with less permissive nucleotides in the positions 4 or 13 of the binding site. We tested this hypothesis by comparing the *in vitro* DNA binding ability of Cmr WT, CmrNΔ12 and Cmr-Arg_{loop}* proteins to DNA sequences containing either a WT or mutated Cmr binding site from PRv1675c. All three proteins bound and shifted the WT DNA sequence (Figures 4B and 5). While Cmr WT and CmrNΔ12 proteins also retained the ability to bind the singly mutated DNA sequences (4C:G or 13G:C), Cmr-Arg_{loop}* failed to bind the mutated sequences. These results show that the Arg-loop interaction is crucial to Cmr's ability to bind DNA when the core binding site contains less permissive nucleotides at positions 4 or 13. We also tested the ability of Cmr WT and mutant proteins to bind the Cmr ($flank_{CRP}$) DNA fragment to control for the contribution of flanking sequences to Cmr-Arg_{loop}*'s inability to bind the mutated sequences. Cmr-Arg_{loop}* bound the Cmr ($flank_{CRP}$) DNA sequence, suggesting that the specificity determinants are contained within the core DNA sequence motif rather than the flanking sequences.

DNA bending by Cmr

CRP_{Ec} and CRP_{Mt} bend DNA upon binding (20,58), so we investigated the ability of Cmr to bend DNA using five identically sized fragments in which the Cmr binding site is positioned at different distances relative to the center (schematic in Figure 6A). Typically, protein–DNA complexes with the binding site most centrally located have the slowest electrophoretic mobility, and the complexes with binding sites farthest away from the center migrate fastest through a non-denaturing gel, if DNA bending occurs at the TF binding site. We noted that identically sized DNA fragments from PRv1675c region migrated slightly differently in the native gel, even in the absence of protein, suggesting some natural bending in this region (Figure 6B).

Surprisingly, Cmr WT caused an atypical DNA migration pattern consistent with an asymmetric bend at the PRv1675c binding site (Figure 6B). Similar migration anomalies were also observed with CmrΔN12 or Cmr-Arg_{loop}* in complex with the DNA fragments from the PRv1675c region. With the exception of DNA fragments D and E, Cmr-Arg_{loop}*–DNA complexes migrated much

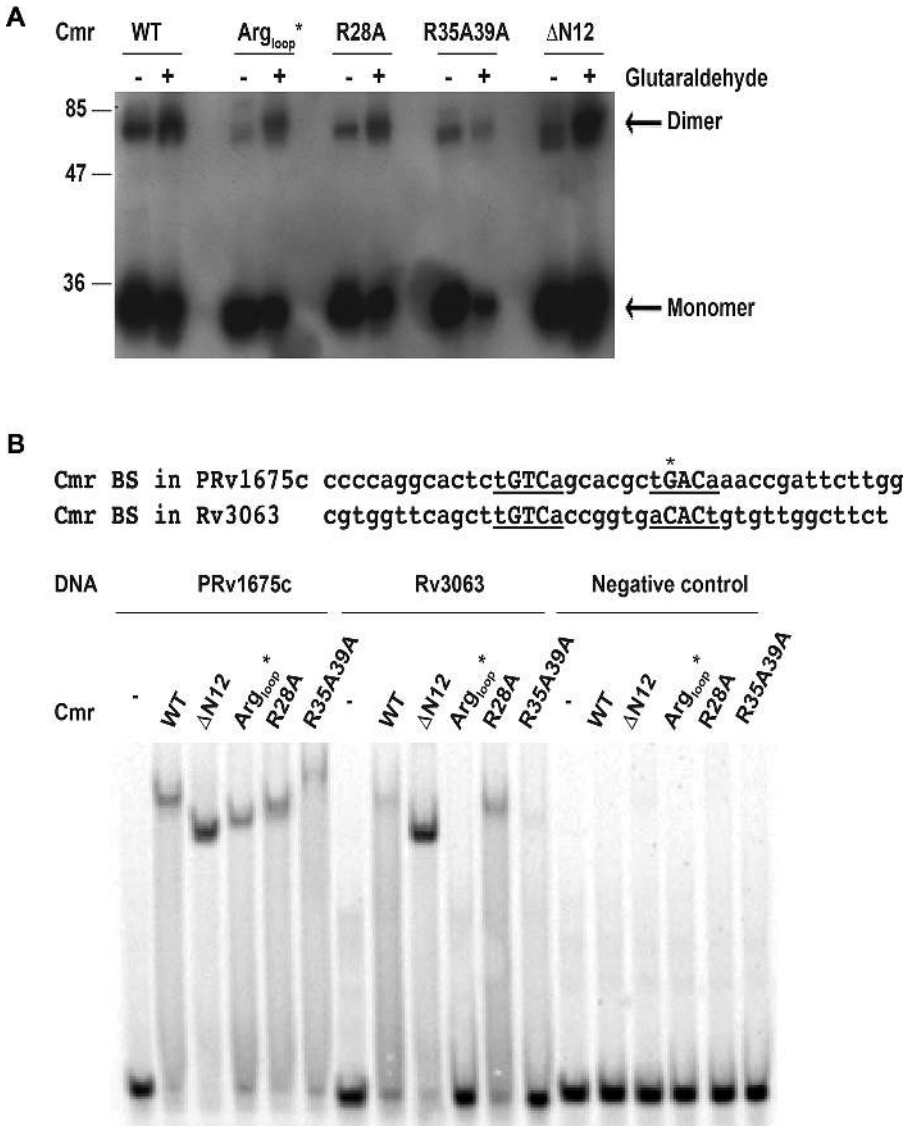

Figure 4. Dimerization and DNA binding by Cmr WT and mutant proteins. (**A**) Glutaraldehyde crosslinking of purified WT and mutant Cmr proteins analyzed by western blotting using Rabbit anti-Cmr antiserum. The molecular weight of Cmr is 26.7 kDa. Note that the overexpressed proteins contain a C-terminal His6x tag of about 1.5 kDa. Therefore, the recombinant protein migrates at about 28.2 kDa. (**B**) EMSA showing DNA binding ability of Cmr WT and mutant proteins to 40-bp dsDNA fragments containing Cmr binding sites from PRv1675c and Rv3063 region. Negative control probe used here is a 40-bp dsDNA from Rv3063 region.

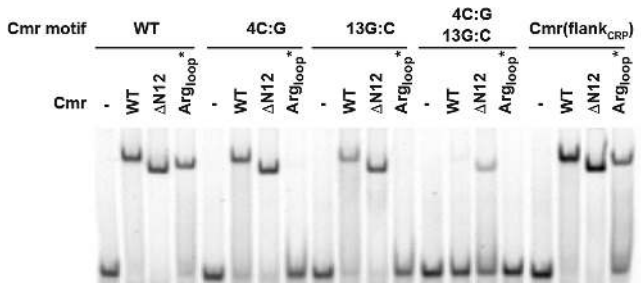

Figure 5. R-loop allows binding to sites with less permissive nucleotides. EMSA showing DNA binding ability of Cmr WT and mutant proteins–CmrΔN12 and Cmr-Arg$_{loop}$* with the WT or mutated Cmr binding site in PRv1675c. Cmr (flank$_{CRP}$) DNA fragment contains the WT 16-bp palindromic Cmr binding site from PRv1675c, but flanking sequences from Rv0885c-serC region.

faster than corresponding Cmr WT–DNA complexes. This suggests that the Arg-loop's interaction with the DNA exerts specific conformational effects on the DNA–protein complex when the binding site is distal to the left end of the DNA probe. The change in charge due to substitution of four arginine residues with alanine in Cmr-Arg$_{loop}$* is unlikely to be the reason for the faster electrophoretic mobility of Cmr-Arg$_{loop}$*–DNA complexes A, B and C, as DNA probes D and E in complex with Cmr WT or Cmr-Arg$_{loop}$* protein migrate similarly through the gel. Rather, altered rigidity and/or shape due to the additional Arg loop contacts may explain this migration anomaly.

We performed a similar experiment with identically sized DNA fragments containing the Cmr binding site in the Rv3063 region (Figure 6C). These DNA fragments showed

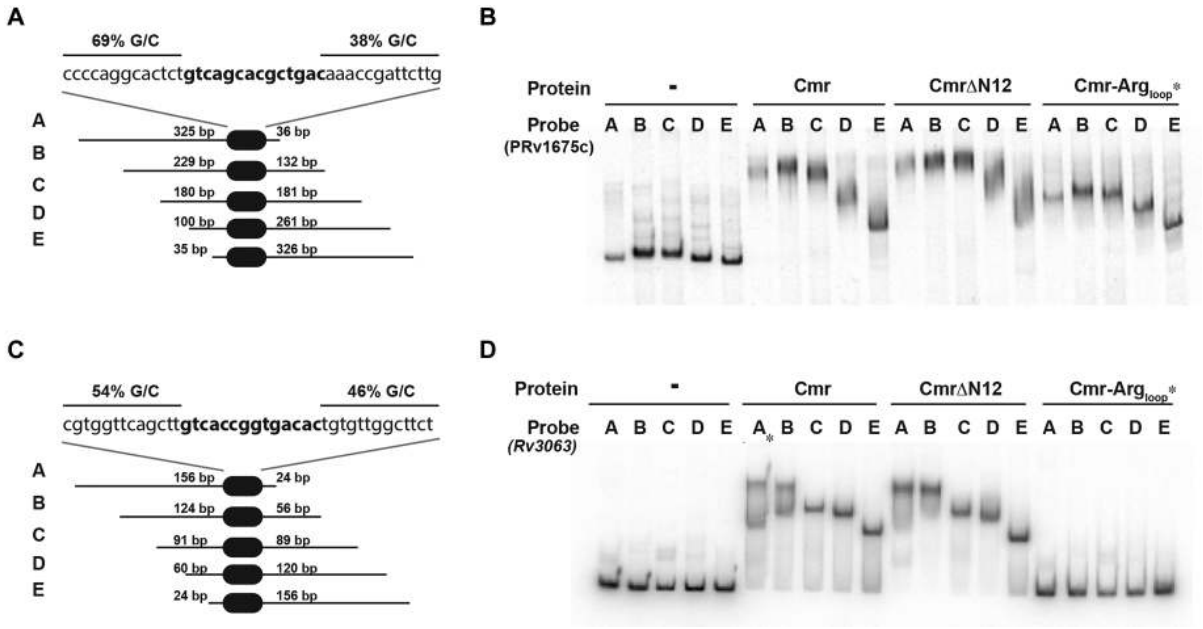

Figure 6. DNA bending by Cmr WT and mutant proteins. (**A**) Graphic showing five equal sized fragments (denoted A to E) with the Cmr binding site from the PRv1675c locus (represented as red ovals) at different positions relative to the center. These fragments were amplified by PCR, and used as DNA probes in EMSA (**B**). EMSA showing the mobility of 401-bp sized fragments A to E, amplified from the PRv1675c locus, in complex with Cmr WT, CmrΔN12 and Cmr-Arg$_{loop}$*. (**C**) Graphic showing five equal sized fragments (denoted A to E) with the Cmr binding site from the intragenic locus in Rv3063 (represented as ovals) at different positions relative to the center. These fragments were amplified by PCR, and used as DNA probes in EMSA (**D**). EMSA showing the mobility of 220-bp sized fragments A to E, amplified from the intragenic locus in Rv3063, in complex with Cmr WT and CmrΔN12. Cmr-Arg$_{loop}$* does not bind this fragment as shown in Figure 4B.

no natural bends in DNA in the absence of Cmr binding (Figure 6D). In the presence of Cmr WT protein, two distinct bands were observed with DNA probes A and B. However, on addition of CmrΔN12, DNA–protein complexes with probes A or B migrated as single bands, similar to what was observed with PRv1675c fragments. As expected, Cmr-Arg$_{loop}$* did not bind these fragments, as this binding site contains the less preferred nucleotide C at motif position 13 (Figure 4B). Together, these results demonstrate that Cmr bends DNA upon binding. The interaction of the Arg-loop likely has a conformational effect on the DNA-protein complex that affects its electrophoretic migration. However the Arg-loop is not required for DNA bending, as the atypical migration pattern observed with Cmr WT–DNA complexes is also observed with Cmr-Arg$_{loop}$*–DNA complexes (Figure 6B). The role of the N-terminal region is less clear, but differences in migration between Rv3063 DNA–protein complexes containing CmrΔN12 versus WT Cmr suggest some impact of the N-terminal region on Cmr's DNA interaction.

CmrNΔ12 binds DNA with higher affinity than Cmr WT

CmrNΔ12 showed stronger binding to the Rv3063 and PRv1675c DNA fragments, whereas Cmr-Arg$_{loop}$* showed weaker binding to the PRv1675c fragment and no binding to Rv3063 DNA fragments, compared to the WT protein (Figures 4B and 5). We determined the binding affinities (K_d) of Cmr WT, CmrNΔ12 and Cmr-Arg$_{loop}$* proteins to the 40-bp dsDNA fragment from PRv1675c (Figure 7). CmrNΔ12 bound DNA with ∼ 3-fold higher affinity than WT protein, whereas Cmr-Arg$_{loop}$* bound DNA with ∼2-fold lower affinity than WT (K_d of Cmr WT = 8.3 ± 0.5 nM; K_d of CmrNΔ12 = 2.7 ± 0.1 nM; K_d of Cmr-Arg$_{loop}$* = 13.3 ± 0.9 nM). We were also interested to note that Cmr WT, CmrNΔ12 and Cmr-Arg$_{loop}$* bound the DNA sequence containing a single Cmr binding site with a Hill coefficient (h) ≥ 2, indicating strong positive cooperativity in DNA binding. Such strong positive cooperativity has not been reported for CRP$_{Ec}$ or CRP$_{Mt}$ (19,30), consistent with our finding of Hill coefficients of 1.2 for CRP$_{Ec}$ and 1.3 for CRP$_{Mt}$ (Supplementary Figure S2).

Model for DNA binding by Cmr using HTH domain and Arg-loop

We propose a model of DNA binding by Cmr where, in addition to the core DNA binding contacts made by the HTH domain, the Arg-loop makes supplementary contacts with the sequences flanking the core-binding site (Figure 8A). Cmr may bend the DNA upon binding in such a way that the determinant nucleotides are positioned to make contacts with the amino acids in the HTH. This bending would also bring the flanking DNA sequences in close proximity to the Arg-loop. It is possible that the HTH mediated binding is sufficient at a binding site that contains all preferred nucleotides, and the Arg-loop contacts serve an accessory purpose. However, these accessory contacts may become critical when the DNA binding site contains less preferred nucleotides, such as 4G or 13C, which cannot be accommodated well by Cmr's HTH.

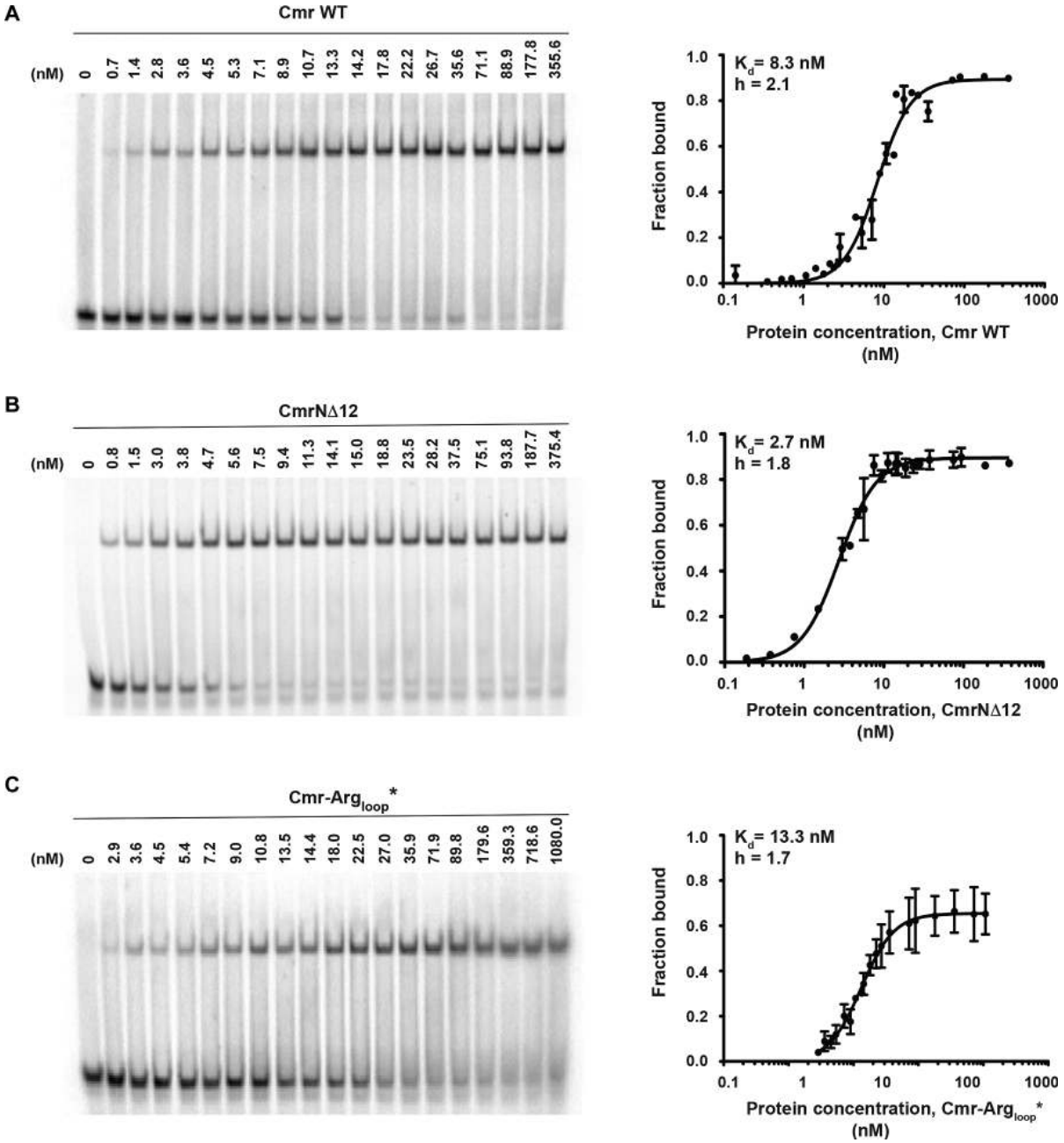

Figure 7. Binding affinity of Cmr WT, CmrΔN12 and Cmr-Arg$_{loop}$*. EMSA using radiolabeled 40-bp WT Cmr binding site from PRv1675c as DNA probe, and titrating increasing concentrations of proteins (**A**) Cmr WT (**B**) CmrΔN12 and (**C**) Cmr-Arg$_{loop}$*. Representative EMSA is shown here. The K_d values reported in the text are average of two biological repeats ± SD.

DISCUSSION

The structure-guided functional analyses in this study provide important new insights into the biology of a TF that contributes to Mtb virulence and pathogenesis. We found that a DNA binding motif derived from Cmr's genomic binding sites does not fully reflect Cmr's high degree of selectivity for DNA binding interactions. Rather, Cmr's DNA binding specificity is determined by multiple levels of DNA interaction, allowing it to recognize subtle differences in DNA binding sites with remarkable specificity. This unexpected complexity likely masks some of Cmr's positional sequence preferences when the DNA binding motif is derived cumulatively from multiple types of binding sites throughout the genome.

These studies also identified novel structural features of Cmr involved in DNA binding in addition to the HTH DNA binding domain characteristic of the CRP/FNR family. These features include an unusual N-terminal domain and an arginine rich loop, not found in prototype family members CRP$_{Ec}$, FNR or CRP$_{Mt}$, distinguish Cmr as a novel member of the CRP/FNR family. The unique fold of Cmr also indicates a more distant evolutionary relationship

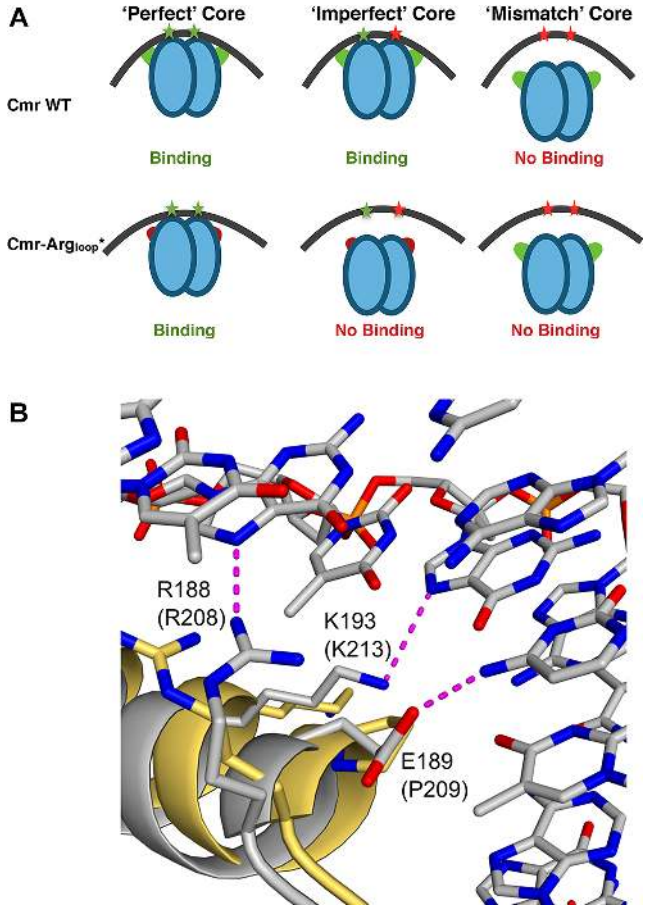

Figure 8. Modeling Cmr–DNA binding interactions. (**A**) Proposed mechanism of DNA binding by Cmr using HTH domain and Arg-loop. The black curved line indicates the DNA; red stars indicate less preferred nucleotide and green stars indicate preferred nucleotides at position 4 or 13 in the binding motif. The arginine loop is denoted by green loops on the Cmr protein schematic (blue ovals), and red semi-circles denote mutated Arg-loops. Productive DNA contacts with either the DNA-binding helices or Arg-loops are indicated schematically. (**B**) Differences within the DNA binding helix of Cmr and CRP_{Mt}. The structures of Cmr (yellow ribbons and carbons) and DNA-bound CRP_{Mt} (gray ribbons and carbons) are superimposed. Helix F of each structure is drawn as a ribbon, and critical side chains of CRP_{Mt} involved in DNA binding are drawn as sticks with hydrogen bonding depicted in magenta. Corresponding side chains of Cmr are also drawn as sticks and labeled. Residue numbering in parentheses refers to structurally equivalent Cmr positions. Nitrogen, oxygen and phosphorus atoms are colored blue, red and orange, respectively.

with CRP/FNR TFs, as suggested previously by a phylogenetic study (6). This functional complexity may provide Cmr with the flexibility to simultaneously integrate multiple signals and modulate its DNA interactions in response to different inputs, including interactions with other TFs, such as those reported for DosR (26,27).

Additional structural features involved in DNA binding

The position of the Arg-loop proximal to the HTH domain is consistent with direct Arg-loop–DNA interaction (Figure 3D), as is the observation that the Arg-loop is needed for Cmr binding at sites that contain a less preferred nucleotide at motif position 4 or 13 (Figures 4B, 5 and 7C). However, it is not known whether the Arg-loop contacts specific nucleotides in the DNA sequences flanking the core-binding site. Analysis of Cmr binding sites from the prior ChIP-seq studies (24,26) failed to identify any additional conserved nucleotides in the sequences that flank the core-binding sites (Supplementary Figure S1a and unpublished studies). Also, replacing the DNA sequences flanking the core-binding site had no significant impact on Cmr's DNA binding specificity in the presence of the preferred discriminator bases (Figure 5). An alternate possibility is that the Arg-loop recognizes local DNA shape. Arginine residues have been associated with DNA-shape based interactions by other TFs, particularly in the minor groove (55). Our results suggest that this additional Arg-loop–DNA interaction becomes important only when one of the DNA binding helices lacks contact with a preferred discriminator base, so future studies will address questions regarding the nature of this contact in the context of different types of core binding sites.

The role of the Cmr N-terminus in DNA binding is not clear. CmrNΔ12 bound DNA with higher affinity than WT and Cmr-Arg_{loop}* proteins (Figure 7B), suggesting that the N-terminal region restricts the DNA binding activity of the full-length protein. Consistent with our structural data, DISOPRED (53) and RONN (52) predict residues 1–19 in the N-terminal region of Cmr to be disordered and potentially involved in protein–protein interaction (data not shown). Intrinsically disordered protein domains are being increasingly identified in eukaryotes, prokaryotes and archaea (59). Recent studies suggest a higher prevalence of disordered proteins in bacteria with high GC genomes (59), such as Mtb. It has been proposed that the structural flexibility associated with disordered regions allows them to adopt different conformations based on their physiological environments, facilitating interactions with diverse protein partners (60,61). We speculate that the N-terminal region of Cmr serves a biochemical regulatory role *in vivo* by modulating Cmr's binding to specific DNA sites in response to physiological signals. Such modulatory signals could include co-factor binding or post-translational modifications that affect binding affinity and/or specificity.

Balancing specificity and plasticity of DNA binding

Cmr failed to bind some CRP_{Mt} binding sites based on the discriminatory nucleotides at positions 4 and 13 of the MEME motif (Figure 2C). This inability of Cmr to bind CRP_{Mt} binding sites could in part be attributed to the differences in the HTH DNA binding domain of the two TFs. Three residues (Arg188, Glu189 and Lys193) in the HTH DNA binding domains of CRP_{Mt} make direct DNA contacts with the critical nucleotides in the core binding motif (Figure 8B). The Glu189 residue on each subunit of homodimeric CRP_{Mt} contacts nucleotide 4G or 13C in the motif DNA sequence (20,21,58). However, Glu189 is replaced by Pro209 in Cmr's HTH domain, which likely results in an unfavorable interaction between the Cmr HTH domain and the DNA probe in which motif positions 4 or 13 are G and C, respectively (Figure 2).

Both the structure and experimental results indicate that the Arg-loop supports the DNA binding function of the HTH domain in Cmr. The Arg-loop–DNA interaction be-

comes particularly critical to the DNA binding ability of Cmr at binding sites that contain less preferred nucleotides at motif position 4 or 13 (Figures 4B and 5). This additional interaction with DNA via the Arg loop also confers plasticity of DNA binding to Cmr, which is more tolerant than CRP_{Mt} to changes at motif positions 4 and 13. Single point mutations at either position negatively affect Cmr's binding, but do not completely eliminate it (Figure 2D). In contrast, mutating positions 4 or 13 causes complete loss of CRP_{Mt} binding to the mutated sequence (Figure 2C). The Arg loop, present only in Cmr and absent from CRP_{Mt}, is required to tolerate changes at these nucleotide positions (Figure 5). Different mechanisms are used by other TFs to distinguish between closely related sequences, or in some cases broaden the binding site repertoire. For example, the Drosophila genome encodes eight Hox proteins that recognize and bind identical binding sites (62,63). However, interaction of these Hox proteins with *in vivo* co-factors confers a 'latent specificity' to the proteins, allowing them to differentiate between sites that vary only at two central nucleotide positions (64,65). In *E. coli*, ligation of a metal cluster by IscR relieves an unfavorable DNA–protein interaction by repositioning a single amino acid residue. This broadens the DNA binding specificity of IscR, allowing it to bind two different types of motifs (66). Together, the HTH and the Arg-loop in Cmr allow for specificity by recognizing single nucleotide changes in the binding site, as well as plasticity, because the additional Arg-loop–DNA interaction allows Cmr to tolerate some changes at the discriminatory positions of the binding site. Further studies will address the extent to which this enhanced core site binding flexibility is counterbalanced by unknown determinants of Arg-loop–DNA interactions that may restrict Cmr's binding profile in response to factors other than the core binding sequence.

CRP_{Ec} and Fnr have shown reciprocal binding and activation of transcription at some promoters (67). The overlap observed between CRP_{Mt} and Cmr binding sites (Figure 1) raises the possibility of co-regulation of some genes by both these TFs. Future studies focusing on the interaction of Cmr with the transcriptional machinery, particularly at promoters where Cmr and CRP_{Mt} binding sites are overlapping, may provide further insights into the mechanism of transcriptional regulation by Cmr and potential co-regulation of genes by Cmr and CRP_{Mt}.

Cmr bends DNA asymmetrically

We found that Cmr binding to its cognate site induces an asymmetrical bend in the DNA (Figure 6), which contrasts with the symmetrical bending caused by CRP_{Mt} (20) and CRP_{Ec} protein (58). Many prokaryotic as well as eukaryotic DNA binding proteins bend DNA upon binding (68–74), but the reason for this asymmetrical bending with Cmr binding is not clear. Proximal sequences (~3–7 bp) immediately surrounding the core-binding site tend to influence TF binding more than the distal sequences (75). Swapping the flanking sequences around the core DNA binding motif reversed the direction of the asymmetry (data not shown), suggesting that the flanking sequences influence Cmr's asymmetric DNA bending pattern. The extent to which G/C content of the DNA immediately adjacent the 16-bp core-binding site contributes to the asymmetric bending of DNA upon Cmr binding is not clear (Figure 6). A DNA probe that contains flanking sequences ~50% G/C showed a pattern of gel mobility with a mixture of symmetrical and asymmetrical DNA bending (Figure 6D), although the symmetrical migration required the entire N-terminal domain. As the nucleotide sequence of a DNA fragment contributes to the DNA topology, it is possible that Cmr recognizes DNA shape motifs in the flanking sequences as an additional method of binding site recognition. Such 'indirect readout' mechanisms that facilitate DNA-proteins interaction have been described previously (76–79).

Cooperative DNA binding suggests a unique mechanism of DNA binding

Cmr WT and mutant proteins displayed strong cooperativity in binding to DNA (Hill coefficient ≈2) (80), whereas DNA binding experiments using CRP_{Mt} and CRP_{Ec} showed much less cooperativity (Hill coefficient of 1.3 and 1.2, respectively) (Figure 7 and Supplementary Figure S2). This divergence in cooperativity may reflect mechanistic differences in DNA binding by Cmr, compared to CRP_{Ec} and CRP_{Mt}, which are thought to dimerize prior to DNA binding (54,58,67,81). In the case of *E. coli* Fnr, only the homodimer is capable of binding DNA, and the monomer-dimer equilibrium is controlled by oxygen availability via the degradation/assembly of an oxygen sensitive [4Fe-4S] cluster (82).

Cooperative binding of homodimeric TFs commonly occurs when binding of the TF to one site positively facilitates the binding of a second dimer to another binding site (83–85). However, the 40-bp DNA probes used in our experiments (Figure 7 and Supplementary Figure S2) each contain only a single binding site (i.e. two-palindromic half sites). Therefore, we propose a model in which binding of monomeric Cmr to a palindromic half-site on the DNA facilitates the binding of the second subunit. In this case, DNA binding would both facilitate and stabilize Cmr dimerization. This model is consistent with a Hill coefficient of ~2, although the Hill coefficient cannot be used to determine stoichiometry. Analytical ultracentrifugation experiments of Cmr (unpublished data) showed that the protein was present in a major and a minor population, with the minor population migrating at a position consistent with it being a higher molecular weight complex than the major population. However, absolute molecular weight calculations were inconclusive; neither peak had a predicted molecular weight that was consistent with an integral number of molecules in the complex, making these measurements unreliable. The presence of 10% glycerol in the buffers, required for protein stability, is likely to be responsible for the difficulty in obtaining accurate molecular weight measurements.

A recent report concluded that Cmr is dimeric based on size exclusion chromatography (27). However, migration of proteins in the size exclusion chromatography column is based on the protein's hydrodynamic size and volume and can be affected by shape of the protein, thereby causing similar sized proteins to elute at different times. This same re-

port (27) also presents EMSA data that are consistent with Cmr binding to DNA with positive cooperativity. Together with our structural and biochemical data suggesting that Cmr has disordered protein domains, it is clear that more definitive methods such as size exclusion chromatography coupled with light scattering or analytical ultracentrifugation under different solution conditions will be required to clearly detect the oligomeric state of Cmr in solution.

The hypothesis that homodimerization of Cmr is facilitated and stabilized by DNA is also supported by our initial observation that a single nucleotide change in one palindromic half-site is tolerated well by Cmr (Figure 2D). It is plausible that a Cmr monomer binds to the preferred half site first, and nucleates the dimerization as well as the interaction of the second monomer to the 'imperfect' half-site, with the additional DNA binding support from the Arg-loop. We are currently investigating this hypothesis, as it sets Cmr apart from its CRP/FNR family counterparts, which are thought to dimerize prior to DNA binding.

Absence of an identifiable cAMP binding pocket

The X-ray crystal structure of Cmr revealed the lack of an obvious cAMP binding pocket in the protein (Figure 3), which is consistent with biochemical studies that failed to detect direct cAMP–Cmr interaction (26,27). The helix N1 in Cmr at least partially blocks the region of Cmr that is analogous to the cAMP binding site in CRP_{Mt}. Additionally, the amino acid residues that coordinate cAMP in CRP_{Mt} are not conserved in Cmr (Figure 3B). Thus, Cmr appears to be better related to other CRP-family TFs like Sdrp and TTHB099 from *T. thermophilus* that also lack an effector molecule-binding pocket, and show no conservation of the nucleotide binding residues (86,87). Nonetheless, cAMP-dependent differences in Cmr–DNA interaction have been observed *in vitro* and *in vivo* (26). The observed differences in electromobility of DNA–protein complexes in the presence of cAMP often occurred when multiple Cmr binding sites were present, so it is possible that a cryptic cAMP-binding pocket is formed upon Cmr multimerization, resulting in the differential electrophoretic mobility of DNA–protein complexes (26). Alternatively, a large conformational change in Cmr upon DNA binding could occur that makes a cAMP-binding pocket accessible, or cAMP might gain access to a binding pocket in Cmr monomers before they dimerize on the DNA. The difference in mobility of DNA–Cmr complexes in the presence of cAMP despite the absence of an obvious cAMP-binding pocket is intriguing, and warrants further investigation.

Dithiol-disulfide linkage between cysteine residues is unlikely

A recent study suggested that Cmr might act as a redox sensor via a dithiol-disulfide linkage between the two cysteine residues C36 and C131 in Cmr (27). However, our structural analysis indicates that interaction between C36 and C131 is highly unlikely, as the two residues are positioned too far apart and only C36 is solvent exposed. The side chain of C36 packs against helix C at the dimerization interface between the two Cmr monomers, so covalent modification of this residue could perturb the structure of the dimer and/or reduce affinity for self-association through changes in protein conformation at the dimer interface. In this regard, we note that mutating residues R35 and R39, which immediately surround the C36 residue, also had negative effects on Cmr's DNA binding ability in our study (Figure 4). We speculate that the non-conserved N-terminal extension of Cmr (residues 1–39) is sensitive to cellular signals, and can modulate DNA binding by altering the protein conformation and/or its ability to dimerize. The N-terminal extension would thus provide a novel regulatory response mechanism for Cmr.

In summary, we have identified novel regions in Cmr that are involved in its DNA binding ability and specificity of recognition. The additional DNA binding regions not only confer specificity but also a plasticity to Cmr's DNA binding. Our data point towards a distinct mode of Cmr binding to DNA, which sets it apart from CRP_{Ec} and CRP_{Mt}, making Cmr a very unusual member of CRP/FNR family of TFs. Future studies are directed towards understanding the role of Cmr in Mtb biology with a focus on mechanisms underlying co-regulation of genes proximal to overlapping Cmr and CRP_{Mt} binding sites.

DATA AVAILABILITY

The atomic coordinates and structure factors for Cmr1–244 and Cmr 13–244 have been deposited into the Protein Data Bank under accession codes 5W5A and 5W5B, respectively. ChIPseq data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (88) and are accessible through GEO Series accession number GSE106135 (https://www.ncbi.nih.gov/geo/query/acc.cgi?acc=GSE106135).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Michael Rudolph at the New York Structural Biology Center for aid in X-ray data collection and Randy Abramowitz for help at synchrotron beamline X4C. Beamline X4C is supported by the New York Structural Biology Center at the National Synchrotron Light Source of Brookhaven National Laboratory, a Department of Energy Facility. We also thank Dr Joseph Wade and Dr Jing Wang at the Wadsworth Center, NY for helpful discussions and for providing purified CRP_{Ec} protein, and Dr Navjot Singh for help with bioinformatics analysis of ChIP-seq data. We acknowledge the Wadsworth Center's Applied Genomics Technologies, Biochemistry, Bioinformatics & Statistics and Media and Tissue Culture Core facilities for technical support and expertise.

FUNDING

National Institutes of Health (NIH) [R01AI063499 to K.A.M., R01GM080573 to J.D.P., in part]. Funding for open access charge: NIH [R01AI063499].

Conflict of interest statement. None declared.

REFERENCES

1. WHO (2016) *Global Tuberculosis Report 2016*. World Health Organization, Geneva.
2. McDonough,K.A. and Rodriguez,A. (2011) The myriad roles of cyclic AMP in microbial pathogens: from signal to sword. *Nat. Rev. Microbiol.*, **10**, 27–38.
3. Bai,G., Knapp,G.S. and McDonough,K.A. (2011) Cyclic AMP signalling in mycobacteria: redirecting the conversation with a common currency. *Cell Microbiol.*, **13**, 349–358.
4. McCue,L.A., McDonough,K.A. and Lawrence,C.E. (2000) Functional classification of cNMP-binding proteins and nucleotide cyclases with implications for novel regulatory pathways in *Mycobacterium tuberculosis*. *Genome Res.*, **10**, 204–219.
5. Shaw,D.J., Rice,D.W. and Guest,J.R. (1983) Homology between CAP and Fnr, a regulator of anaerobic respiration in *Escherichia coli*. *J. Mol. Biol.*, **166**, 241–247.
6. Korner,H., Sofia,H.J. and Zumft,W.G. (2003) Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol. Rev.*, **27**, 559–592.
7. Kolb,A., Busby,S., Buc,H., Garges,S. and Adhya,S. (1993) Transcriptional regulation by cAMP and its receptor protein. *Annu. Rev. Biochem.*, **62**, 749–795.
8. Matsui,M., Tomita,M. and Kanai,A. (2013) Comprehensive computational analysis of bacterial CRP/FNR superfamily and its target motifs reveals stepwise evolution of transcriptional networks. *Genome Biol. Evol.*, **5**, 267–282.
9. Reynolds,M.F., Parks,R.B., Burstyn,J.N., Shelver,D., Thorsteinsson,M.V., Kerby,R.L., Roberts,G.P., Vogel,K.M. and Spiro,T.G. (2000) Electronic absorption, EPR, and resonance raman spectroscopy of CooA, a CO-sensing transcription activator from *R. rubrum*, reveals a five-coordinate NO-heme. *Biochemistry*, **39**, 388–396.
10. Hu,P., Leighton,T., Ishkhanova,G. and Kustu,S. (1999) Sensing of nitrogen limitation by Bacillus subtilis: comparison to enteric bacteria. *J. Bacteriol.*, **181**, 5042–5050.
11. Ikeda,T.P., Shauger,A.E. and Kustu,S. (1996) *Salmonella typhimurium* apparently perceives external nitrogen limitation as internal glutamine limitation. *J. Mol. Biol.*, **259**, 589–607.
12. McKay,D.B. and Steitz,T.A. (1981) Structure of catabolite gene activator protein at 2.9 A resolution suggests binding to left-handed B-DNA. *Nature*, **290**, 744–749.
13. Nakajima,H., Honma,Y., Tawara,T., Kato,T., Park,S.Y., Miyatake,H., Shiro,Y. and Aono,S. (2001) Redox properties and coordination structure of the heme in the co-sensing transcriptional activator CooA. *J. Biol. Chem.*, **276**, 7055–7061.
14. Busby,S. and Ebright,R.H. (1999) Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.*, **293**, 199–213.
15. Green,J., Scott,C. and Guest,J.R. (2001) Functional versatility in the CRP-FNR superfamily of transcription factors: FNR and FLP. *Adv. Microb. Physiol.*, **44**, 1–34.
16. Klinger,A., Schirawski,J., Glaser,P. and Unden,G. (1998) The *fnr* gene of *Bacillus licheniformis* and the cysteine ligands of the C-terminal FeS cluster. *J. Bacteriol.*, **180**, 3483–3485.
17. Lampidis,R., Gross,R., Sokolovic,Z., Goebel,W. and Kreft,J. (1994) The virulence regulator protein of *Listeria ivanovii* is highly homologous to PrfA from *Listeria monocytogenes* and both belong to the Crp-Fnr family of transcription regulators. *Mol. Microbiol.*, **13**, 141–151.
18. Vega,Y., Dickneite,C., Ripio,M.T., Bockmann,R., Gonzalez-Zorn,B., Novella,S., Dominguez-Bernal,G., Goebel,W. and Vazquez-Boland,J.A. (1998) Functional similarities between the *Listeria monocytogenes* virulence regulator PrfA and cyclic AMP receptor protein: the PrfA* (Gly145Ser) mutation increases binding affinity for target DNA. *J. Bacteriol.*, **180**, 6655–6660.
19. Bai,G., Gazdik,M.A., Schaak,D.D. and McDonough,K.A. (2007) The *Mycobacterium bovis* BCG cyclic AMP receptor-like protein is a functional DNA binding protein in vitro and in vivo, but its activity differs from that of its *M. tuberculosis* ortholog, Rv3676. *Infect. Immun.*, **75**, 5509–5517.
20. Bai,G., McCue,L.A. and McDonough,K.A. (2005) Characterization of *Mycobacterium tuberculosis* Rv3676 (CRP_{Mt}), a cyclic AMP receptor protein-like DNA binding protein. *J. Bacteriol.*, **187**, 7795–7804.
21. Rickman,L., Scott,C., Hunt,D.M., Hutchinson,T., Menéndez,M.C., Walan,R., Hinds,J., Colston,M.J., Green,J. and Buxton,R.S. (2005) A member of the cAMP receptor protein family of transcription regulators in *Mycobacterium tuberculosis* is required for virulence in mice and controls transcription of the *rpfA* gene coding for a resuscitation promoting factor. *Mol. Microbiol.*, **56**, 1274–1286.
22. Stapleton,M., Haq,I., Hunt,D.M., Arnvig,K.B., Artymiuk,P.J., Buxton,R.S. and Green,J. (2010) Mycobacterium tuberculosis cAMP receptor protein (Rv3676) differs from the Escherichia coli paradigm in its cAMP binding and DNA binding properties and transcription activation properties. *J. Biol. Chem.*, **285**, 7016–7027.
23. Agarwal,N., Raghunand,T.R. and Bishai,W.R. (2006) Regulation of the expression of whiB1 in Mycobacterium tuberculosis: role of cAMP receptor protein. *Microbiology*, **152**, 2749–2756.
24. Knapp,G.S., Lyubetskaya,A., Peterson,M.W., Gomes,A.L., Ma,Z., Galagan,J.E. and McDonough,K.A. (2015) Role of intragenic binding of cAMP responsive protein (CRP) in regulation of the succinate dehydrogenase genes Rv0249c-Rv0247c in TB complex mycobacteria. *Nucleic Acids Res.*, **43**, 5377–5393.
25. Kahramanoglou,C., Cortes,T., Matange,N., Hunt,D.M., Visweswariah,S.S., Young,D.B. and Buxton,R.S. (2014) Genomic mapping of cAMP receptor protein (CRP_{Mt}) in *Mycobacterium tuberculosis*: relation to transcriptional start sites and the role of CRPMt as a transcription factor. *Nucleic Acids Res.*, **42**, 8320–8329.
26. Ranganathan,S., Bai,G., Lyubetskaya,A., Knapp,G.S., Peterson,M.W., Gazdik,M., AL,C.G., Galagan,J.E. and McDonough,K.A. (2016) Characterization of a cAMP responsive transcription factor, Cmr (Rv1675c), in TB complex mycobacteria reveals overlap with the DosR (DevR) dormancy regulon. *Nucleic Acids Res.*, **44**, 134–151.
27. Smith,L.J., Bochkareva,A., Rolfe,M.D., Hunt,D.M., Kahramanoglou,C., Braun,Y., Rodgers,A., Blockley,A., Coade,S., Lougheed,K.E.A. *et al.* (2017) Cmr is a redox-responsive regulator of DosR that contributes to *M. tuberculosis* virulence. *Nucleic Acids Res.*, **45**, 6600–6612.
28. Bai,G., Schaak,D.D., Smith,E.A. and McDonough,K.A. (2011) Dysregulation of serine biosynthesis contributes to the growth defect of a *Mycobacterium tuberculosis crp* mutant. *Mol. Microbiol.*, **82**, 180–198.
29. Gazdik,M.A., Bai,G., Wu,Y. and McDonough,K.A. (2009) Rv1675c (cmr) regulates intramacrophage and cyclic AMP-induced gene expression in *Mycobacterium tuberculosis*-complex mycobacteria. *Mol. Microbiol.*, **71**, 434–448.
30. Ebright,R.H., Ebright,Y.W. and Gunasekera,A. (1989) Consensus DNA site for the *Escherichia coli* catabolite gene activator protein (CAP): CAP exhibits a 450-fold higher affinity for the consensus DNA site than for the *E. coli lac* DNA site. *Nucleic Acids Res.*, **17**, 10295–10305.
31. Harman,J.G. (2001) Allosteric regulation of the cAMP receptor protein. *Biochim. Biophys. Acta*, **1547**, 1–17.
32. Donnelly,M.I., Zhou,M., Millard,C.S., Clancy,S., Stols,L., Eschenfeldt,W.H., Collart,F.R. and Joachimiak,A. (2006) An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. *Protein Expr. Purif.*, **47**, 446–454.
33. Aslanidis,C. and de Jong,P.J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.*, **18**, 6069–6074.
34. Studier,F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.*, **41**, 207–234.
35. Cheung,J., Ginter,C., Cassidy,M., Franklin,M.C., Rudolph,M.J., Robine,N., Darnell,R.B. and Hendrickson,W.A. (2015) Structural insights into mis-regulation of protein kinase A in human tumors. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 1374–1379.
36. Blommel,P.G. and Fox,B.G. (2007) A combined approach to improving large-scale production of tobacco etch virus protease. *Protein Expr. Purif.*, **55**, 53–68.
37. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G., McCoy,A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.
38. Wang,J.W., Chen,J.R., Gu,Y.X., Zheng,C.D., Jiang,F., Fan,H.F., Terwilliger,T.C. and Hao,Q. (2004) SAD phasing by combination of

direct methods with the SOLVE/RESOLVE procedure. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 1244–1253.
39. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
40. 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.
41. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
42. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
43. Chen,V.B., Arendall,W.B. 3rd, 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. D Biol. Crystallogr.*, **66**, 12–21.
44. Bailey,T.L., Boden,M., Buske,F.A., Frith,M., Grant,C.E., Clementi,L., Ren,J., Li,W.W. and Noble,W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
45. Afek,A., Schipper,J.L., Horton,J., Gordan,R. and Lukatsky,D.B. (2014) Protein-DNA binding in the absence of specific base-pair recognition. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 17140–17145.
46. Gordan,R., Shen,N., Dror,I., Zhou,T., Horton,J., Rohs,R. and Bulyk,M.L. (2013) Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep.*, **3**, 1093–1104.
47. Stringham,J.L., Brown,A.S., Drewell,R.A. and Dresch,J.M. (2013) Flanking sequence context-dependent transcription factor binding in early Drosophila development. *BMC Bioinformatics*, **14**, 298–310.
48. Holm,L. and Rosenstrom,P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.*, **38**, W545–W549.
49. Townsend,P.D., Jungwirth,B., Pojer,F., Bussmann,M., Money,V.A., Cole,S.T., Puhler,A., Tauch,A., Bott,M., Cann,M.J. *et al.* (2014) The crystal structures of apo and cAMP-bound GlxR from *Corynebacterium glutamicum* reveal structural and dynamic changes upon cAMP binding in CRP/FNR family transcription factors. *PLoS One*, **9**, e113265.
50. Kumar,P., Joshi,D.C., Akif,M., Akhter,Y., Hasnain,S.E. and Mande,S.C. (2010) Mapping conformational transitions in cyclic AMP receptor protein: crystal structure and normal-mode analysis of *Mycobacterium tuberculosis* apo-cAMP receptor protein. *Biophys. J.*, **98**, 305–314.
51. Robert,X. and Gouet,P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.*, **42**, W320–W324.
52. Yang,Z.R., Thomson,R., McNeil,P. and Esnouf,R.M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, **21**, 3369–3376.
53. Ward,J.J., McGuffin,L.J., Bryson,K., Buxton,B.F. and Jones,D.T. (2004) The DISOPRED server for the prediction of protein disorder. *Bioinformatics*, **20**, 2138–2139.
54. Reddy,M.C., Palaninathan,S.K., Bruning,J.B., Thurman,C., Smith,D. and Sacchettini,J.C. (2009) Structural insights into the mechanism of the allosteric transitions of *Mycobacterium tuberculosis* cAMP receptor protein. *J. Biol. Chem.*, **284**, 36581–36591.
55. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.
56. Chen,B., Himes,P., Liu,Y., Zhang,Y., Lu,Z., Liu,A., Yan,H. and Kroos,L. (2014) Structure of bacterial transcription factor SpoIIID and evidence for a novel mode of DNA binding. *J. Bacteriol.*, **196**, 2131–2142.
57. Minezaki,Y., Homma,K., Kinjo,A.R. and Nishikawa,K. (2006) Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J. Mol. Biol.*, **359**, 1137–1149.
58. Schultz,S.C., Shields,G.C. and Steitz,T.A. (1991) Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science*, **253**, 1001–1007.
59. Pavlovic-Lazetic,G.M., Mitic,N.S., Kovacevic,J.J., Obradovic,Z., Malkov,S.N. and Beljanski,M.V. (2011) Bioinformatics analysis of disordered proteins in prokaryotes. *BMC Bioinformatics*, **12**, 66–87.
60. Garza,A.S., Ahmad,N. and Kumar,R. (2009) Role of intrinsically disordered protein regions/domains in transcriptional regulation. *Life Sci.*, **84**, 189–193.
61. Dyson,H.J. and Wright,P.E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.*, **6**, 197–208.
62. Berger,M.F., Badis,G., Gehrke,A.R., Talukder,S., Philippakis,A.A., Pena-Castillo,L., Alleyne,T.M., Mnaimneh,S., Botvinnik,O.B., Chan,E.T. *et al.* (2008) Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell*, **133**, 1266–1276.
63. Noyes,M.B., Christensen,R.G., Wakabayashi,A., Stormo,G.D., Brodsky,M.H. and Wolfe,S.A. (2008) Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell*, **133**, 1277–1289.
64. Crocker,J., Abe,N., Rinaldi,L., McGregor,A.P., Frankel,N., Wang,S., Alsawadi,A., Valenti,P., Plaza,S., Payre,F. *et al.* (2015) Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell*, **160**, 191–203.
65. Amin,S., Donaldson,I.J., Zannino,D.A., Hensman,J., Rattray,M., Losa,M., Spitz,F., Ladam,F., Sagerstrom,C. and Bobola,N. (2015) Hoxa2 selectively enhances Meis binding to change a branchial arch ground state. *Dev. Cell*, **32**, 265–277.
66. Rajagopalan,S., Teter,S.J., Zwart,P.H., Brennan,R.G., Phillips,K.J. and Kiley,P.J. (2013) Studies of IscR reveal a unique mechanism for metal-dependent regulation of DNA binding specificity. *Nat. Struct. Mol. Biol.*, **20**, 740–747.
67. Sawers,G., Kaiser,M., Sirko,A. and Freundlich,M. (1997) Transcriptional activation by FNR and CRP: reciprocity of binding-site recognition. *Mol. Microbiol.*, **23**, 835–845.
68. Robertson,C.A. and Nash,H.A. (1988) Bending of the bacteriophage lambda attachment site by Escherichia coli integration host factor. *J. Biol. Chem.*, **263**, 3554–3557.
69. Parekh,B.S. and Hatfield,G.W. (1996) Transcriptional activation by protein-induced DNA bending: evidence for a DNA structural transmission model. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 1173–1177.
70. Zinkel,S.S. and Crothers,D.M. (1991) Catabolite activator protein-induced DNA bending in transcription initiation. *J. Mol. Biol.*, **219**, 201–215.
71. Gartenberg,M.R. and Crothers,D.M. (1988) DNA sequence determinants of CAP-induced bending and protein binding affinity. *Nature*, **333**, 824–829.
72. Perez-Martin,J. and Espinosa,M. (1993) Protein-induced bending as a transcriptional switch. *Science*, **260**, 805–807.
73. Onate,S.A., Prendergast,P., Wagner,J.P., Nissen,M., Reeves,R., Pettijohn,D.E. and Edwards,D.P. (1994) The DNA-bending protein HMG-1 enhances progesterone receptor binding to its target DNA sequences. *Mol. Cell. Biol.*, **14**, 3376–3391.
74. Kerppola,T.K. and Curran,T. (1991) Fos-Jun heterodimers and Jun homodimers bend DNA in opposite orientations: implications for transcription factor cooperativity. *Cell*, **66**, 317–326.
75. Levo,M., Zalckvar,E., Sharon,E., Dantas Machado,A.C., Kalma,Y., Lotam-Pompan,M., Weinberger,A., Yakhini,Z., Rohs,R. and Segal,E. (2015) Unraveling determinants of transcription factor binding outside the core binding site. *Genome Res.*, **25**, 1018–1029.
76. Otwinowski,Z., Schevitz,R.W., Zhang,R.G., Lawson,C.L., Joachimiak,A., Marmorstein,R.Q., Luisi,B.F. and Sigler,P.B. (1988) Crystal structure of trp repressor/operator complex at atomic resolution. *Nature*, **335**, 321–329.
77. Hizver,J., Rozenberg,H., Frolow,F., Rabinovich,D. and Shakked,Z. (2001) DNA bending by an adenine–thymine tract and its role in gene regulation. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 8490–8495.
78. Rohs,R., Sklenar,H. and Shakked,Z. (2005) Structural and energetic origins of sequence-specific DNA bending: Monte Carlo simulations of papillomavirus E2-DNA binding sites. *Structure*, **13**, 1499–1509.
79. Hegde,R.S., Grossman,S.R., Laimins,L.A. and Sigler,P.B. (1992) Crystal structure at 1.7 A of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature*, **359**, 505–512.

80. Hill,A.V. (1910) The possible effects of the aggregation of the molecules of haemoglobin on its disassociation curves. *J. Physiol.*, **40**, iv–vii.
81. Gallagher,D.T., Smith,N., Kim,S.K., Robinson,H. and Reddy,P.T. (2009) Profound asymmetry in the structure of the cAMP-free cAMP Receptor Protein (CRP) from *Mycobacterium tuberculosis*. *J. Biol. Chem.*, **284**, 8228–8232.
82. Volbeda,A., Darnault,C., Renoux,O., Nicolet,Y. and Fontecilla-Camps,J.C. (2015) The crystal structure of the global anaerobic transcriptional regulator FNR explains its extremely fine-tuned monomer-dimer equilibrium. *Sci. Adv.*, **1**, e1501086.
83. Ptashne,M., Jeffrey,A., Johnson,A.D., Maurer,R., Meyer,B.J., Pabo,C.O., Roberts,T.M. and Sauer,R.T. (1980) How the lambda repressor and cro work. *Cell*, **19**, 1–11.
84. Zeiser,S., Liebscher,H.V., Tiedemann,H., Rubio-Aliaga,I., Przemeck,G.K., de Angelis,M.H. and Winkler,G. (2006) Number of active transcription factor binding sites is essential for the Hes7 oscillator. *Theor. Biol. Med. Model.*, **3**, 11.
85. Fekete,R.A. and Frost,L.S. (2002) Characterizing the DNA contacts and cooperative binding of F plasmid TraM to its cognate sites at oriT. *J. Biol. Chem.*, **277**, 16705–16711.
86. Agari,Y., Kuramitsu,S. and Shinkai,A. (2012) X-ray crystal structure of TTHB099, a CRP/FNR superfamily transcriptional regulator from *Thermus thermophilus* HB8, reveals a DNA-binding protein with no required allosteric effector molecule. *Proteins*, **80**, 1490–1494.
87. Agari,Y., Kashihara,A., Yokoyama,S., Kuramitsu,S. and Shinkai,A. (2008) Global gene expression mediated by Thermus thermophilus SdrP, a CRP/FNR family transcriptional regulator. *Mol. Microbiol.*, **70**, 60–75.
88. Edgar,R., Domrachev,M. and Lash,A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.