5308–5318 *Nucleic Acids Research, 2018, Vol. 46, No. 10* *Published online 30 April 2018*
doi: 10.1093/nar/gky292

The phage T4 MotA transcription factor contains a novel DNA binding motif that specifically recognizes modified DNA

Maxime G. Cuypers[1,†], Rosanna M. Robertson[1,†], Leslie Knipling[2], M. Brett Waddell[1], Kyung Moon[2], Deborah M. Hinton[2,*] and Stephen W. White[1,*]

[1]Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA and [2]Gene Expression and Regulation Section, Laboratory of Cell and Molecular Biology, NIDDK, National Institutes of Health, Bethesda, MD 20892, USA

Received July 17, 2017; Revised April 03, 2018; Editorial Decision April 04, 2018; Accepted April 12, 2018

ABSTRACT

During infection, bacteriophage T4 produces the MotA transcription factor that redirects the host RNA polymerase to the expression of T4 middle genes. The C-terminal 'double-wing' domain of MotA binds specifically to the MotA box motif of middle T4 promoters. We report the crystal structure of this complex, which reveals a new mode of protein-DNA interaction. The domain binds DNA mostly via interactions with the DNA backbone, but the binding is enhanced in the specific cognate structure by additional interactions with the MotA box motif in both the major and minor grooves. The linker connecting the two MotA domains plays a key role in stabilizing the complex via minor groove interactions. The structure is consistent with our previous model derived from chemical cleavage experiments using the entire transcription complex. α- and β-D-glucosyl-5-hydroxymethyl-deoxycytosine replace cytosine in T4 DNA, and docking simulations indicate that a cavity in the cognate structure can accommodate the modified cytosine. Binding studies confirm that the modification significantly enhances the binding affinity of MotA for the DNA. Consequently, our work reveals how a DNA modification can extend the uniqueness of small DNA motifs to facilitate the specificity of protein-DNA interactions.

INTRODUCTION

Regulation of gene expression is a crucial function in all organisms to ensure that the right genes are expressed at the appropriate time. This control is often exerted at the initiation of transcription, the first step in gene expression. All multi-subunit RNA polymerases (RNAPs) share structural similarities and catalyze RNA synthesis using the same fundamental mechanism. It is therefore not surprising that they share common strategies for transcriptional regulation (1,2). One such strategy is the presence of factor(s), which interact with RNAP and with the DNA. These interactions serve to transform a subpar recognition site (promoter) for RNAP into an excellent site, either by increasing the affinity of the polymerase for the promoter, increasing the rate of DNA unwinding at the start site, and/or increasing the rate at which RNAP clears the promoter (3).

In bacteria, the core RNAP associates with a σ specificity factor to generate the functional holoenzyme (1,2,4,5). The primary σ factor in *Escherichia coli* is σ^{70}, which functions during exponential growth, while alternate σ's are used during other conditions such as times of stress (6,7). Promoter recognition by σ^{70} is achieved through specific interactions between two regions of σ and two DNA elements: Region 2.4/3 with the –10 /extended –10 element and Region 4 with the –35 element (6–8). When bacteriophage T4 infects *E. coli*, the phage hijacks and uses the host RNAP, and it does this in three temporal stages (reviewed in (9)). 'Early' T4 promoters simply compete successfully with host promoters for polymerase, while 'late' promoters employ a specific T4-encoded σ factor that replaces σ^{70}. In contrast, the activation of 'middle' promoters involves two phage proteins, the co-activator AsiA and the activator MotA, that structurally adapt the host RNAP/σ factor complex to specifically recognize the T4 middle promoter sequence. In this novel process called σ appropriation, AsiA binds tightly to free σ^{70} and remodels the structure of Region 4 (10,11) such that this portion of σ^{70} no longer interacts with its partner in RNAP core or with the host –35 element (4,12). Instead,

*To whom correspondence should be addressed. Tel: +1 901 595 1503; Fax: +1 901 595 3032; Email: stephen.white@stjude.org
Correspondence may also be addressed to Deborah M. Hinton. Tel: +1 301 496 9885; Fax: +1 301 402 0053; Email: dhinton@helix.nih.gov
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Present address: Rosanna M. Robertson, US Department of Homeland Security, Science and Technology Directorate, Washington, DC 20528, USA.

the C-terminal residues of σ^{70} become the binding partner for MotA that, in turn, recognizes and binds to the MotA box motif [5′ (a/t)(a/t)(a/t)TGCTTtA3′, centered at position -30] present within T4 middle promoter DNA.

MotA contains two domains connected by a flexible linker region (13). The N-terminal activation domain $MotA^{NTD}$ interacts with the remodeled σ^{70} factor (14), and the C-terminal domain $MotA^{CTD}$ can recognize and bind the MotA box motif (15,16). The structures of $MotA^{NTD}$ (17,18) and $MotA^{CTD}$ in the absence of DNA (19,20) have both been determined. $MotA^{CTD}$ adopts an unusual 'double-wing' saddle fold comprising three α-helices atop an antiparallel 6-stranded β-sheet, and the many basic residues on the exposed surface of the β-sheet saddle suggested that this engages the DNA. This mode of binding was supported by analyses using the chemical cleaving agent FeBABE that generated a model for how MotA binds DNA within the AsiA/RNAP/MotA/DNA transcription complex (21). This model predicts that $MotA^{CTD}$ straddles the major groove. However, it also suggests that the MotA linker contributes to DNA binding by engaging the minor groove in the upstream portion of the MotA motif.

The uniqueness of the double-wing domain suggested that it represents a new way in which a small protein module can recognize and bind DNA. In this paper, we report the crystal structure of the $MotA^{linker/CTD}$–DNA complex as revealed by four independent copies of the complex in the crystal, two cognate and two non-cognate with respect to the MotA box motif. Consistent with our previous model (21), $MotA^{CTD}$ binds the DNA across the major groove via multiple interactions with the flanking sugar-phosphate backbones. It uses just three amino acids to 'read' the MotA box motif, and one of these amino acids is provided by the flexible N-terminal linker region that interacts both specifically and non-specifically with the minor groove. In wild type T4 DNA, position 5 of cytosines are modified via hydroxymethylation/glucosylation. $MotA^{CTD}$ provides a cavity within the complex that is ideally positioned and structured to accommodate this moiety and thereby increase the strength and specificity of the interaction. Binding and mutagenesis experiments based on the crystal structure confirm the importance of this cavity and reveal that the sugar provides an additional recognition determinant for the protein.

MATERIALS AND METHODS

Cloning and protein production

The gene for the C-terminal domain and linker region of MotA ($MotA^{linker/CTD}$, amino acids 93–211) was cloned into the bacterial expression vector pET21a(+) (Novagen), and $MotA^{linker/CTD}$ was expressed in BL21(DE3) cells (22). Cells were grown in 1 L of LB media containing 100 μg/ml carbenicillin at 37°C to an OD_{600} of 0.6. The cells were then induced with isopropyl-β-D-1-thiogalactopyranoside (IPTG, 0.3 mM final concentration) and grown overnight at 18°C. Cells were harvested by centrifugation at 3000 × *g* for 25 min and resuspended in lysis buffer (20 mM Tris–HCl, pH 7.9, 50 mM NaCl, 1 mM β-mercaptoethanol, and 1 mM benzamidine) containing EDTA-free cOmplete protease inhibitor cocktail (Roche). Cells were lysed by sonication, and the cell lysate was centrifuged at 30 000 × *g* for 20 min to remove the cell debris. Poly-ethyleneimine was added to a final concentration of 1%, and DNA was pelleted by centrifugation at 12 000 × *g* for 15 min. Ammonium sulfate was added to the supernatant to 80% saturation, and the precipitate was pelleted by centrifugation at 20 000 × *g* for 15 min. The supernatant was dialyzed against buffer A (50 mM potassium phosphate, pH 6.5, 1 mM EDTA, 1 mM DTT, 1 mM benzamidine, 1 mM PMSF) and loaded onto an S-sepharose ion exchange column, and almost pure protein was eluted using a 0–0.8 M NaCl gradient in buffer A. Final purification was achieved using a gel filtration column equilibrated with buffer B (10 mM potassium phosphate, pH 7.5, 50 mM NaCl, 1 mM β-mercaptoethanol, and 0.02 M EDTA). The pure protein was concentrated to 20 mg/ml using Amicon Ultra 10-NMLW concentrators (Millipore).

Full length MotA used for *in vitro* transcriptions, EMSA and SPR was purified as described (21) by phosphocellulose chromatography, followed by HiTrap SP HP cation exchange chromatography.

Plasmids expressing mutant *motA* were constructed using the Q5 site-directed mutagenesis kit (New England Biolabs) and protocol except that the incubation time of the Kinase-Ligase-DpnI (KLD) reaction was extended to 30 min. Primers (Supplementary Table S1) were designed using NEBaseChanger v1.2.6 (http://nebasechanger.neb.com/) to introduce each *motA* mutation into pNW143 (14), which harbors kanamycin resistance and the sequence of WT *motA* under the inducible arabinose promoter (P_{BAD}), resulting in the plasmids pGFK3034 (*motA*Y134A), pGFK3035 (*motA*R135A), and pGFK3036 (*motA*E153A). pGFK3034 was then used as a template to obtain pGFK3037, which contains the double amino acids changes, *motA*Y134A/R135A. Sanger sequencing analysis by Macrogen (Rockville, MD, USA) verified each mutant.

To isolate the full length MotA mutant proteins, BL21(DE3) cells, containing the appropriate *motA* expression plasmid, were grown at 37°C with shaking in 500 ml of LB media with 40 ug/ml kanamycin to an OD_{600} of 0.4. Expression of *motA* was induced with IPTG (0.2% final concentration), and cells were grown for an additional 2 h at 37°C with shaking. All further steps were carried out at 4°C. Cells were harvested by centrifugation at 13 000 × *g* for 10 min, and pellets (1.5–2.6 g) were resuspended in 42 ml sonication buffer [20 mM Tris–HCl (pH 7.9), 1 mM EDTA, 10% glycerol, 50 mM NaCl, 1 mM ß-mercaptoethanol, 1 mM benzamidine]. After lysis by sonication, the resulting fraction was centrifuged at 8750 × *g* for 30 min to remove cell debris. The supernatant was checked for equivalent conductivity with sonication buffer and corrected as needed. Eight milliliter of an equilibrated 50% slurry of phosphocellulose resin (P11 Cellulose Phosphate cation exchanger, Whatman) was added to the supernatant and gently rocked overnight, then loaded into a 30 ml disposable column with a gravity drip. The column was washed with sonication buffer (20 ml) and then with sonication buffer containing increasing concentrations of NaCl (10 ml of 0.2 M, 10 ml of 0.3 M, 8 × 4 ml of 0.4 M, 10 ml of 0.5 M, and 10 ml of 0.8 M NaCl). Protein was eluted with sonication buffer containing 0.4 M NaCl, pooled, and then concentrated and buffer

exchanged with MotA Binding buffer [50 mM potassium phosphate (pH 6.5), 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 1 mM benzamidine] in an Amicon Ultra-15 centrifugal filter, 10,000 MWCO. The fraction (∼1 ml) was loaded onto a cation exchange column (HiTrap SP HP 1 ml, GE Healthcare), and protein was eluted using a 0–0.5 M NaCl gradient (40 ml) in MotA Binding Buffer. Fractions containing MotA (peak eluting at ∼0.35 NaCl) were pooled, concentrated, and buffer exchanged with Mot GC Buffer [200 mM potassium phosphate (pH 6.5), 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 50% glycerol] and stored at −80°C. Because MotAE153A did not exhibit the same elution profile as the other proteins, the following changes were made for the HiTrap SP HP chromatography: the MotA Binding Buffer contained 50 mM Tris–HCl (pH 8.5) instead of 50 mM potassium phosphate (pH 6.5), and the protein was eluted with a 0–0.7 M NaCl gradient (40 ml), with the peak eluting at ∼0.67 M NaCl.

Core RNAP was purchased from New England Biolabs, and σ^{70} (23) and AsiA (24) were purified as described.

DNA

For crystallization, complimentary 22-base pair DNA oligonucleotides containing the consensus MotA box motif (bolded) were purchased from Integrated DNA Technologies (5′-GAAG**CTTTGCTTAA**TAATCCAC-3′), dissolved in TE buffer at a concentration of 20 mg/ml and annealed by mixing equal volumes in Tris–EDTA (TE) buffer to a final concentration of 2.5 mg/ml. The mixture was placed in a water bath at 100°C for 5 min and then slow-cooled in the water bath to room temperature.

Oligomers used for EMSAs and SPR (unmodified or cytosine 5-hydroxymethyl) were purchased from either e-oligos (gel purified) or Integrated DNA Technologies (HPLC purified). Single-stranded oligomers @ 5 pmol/μl in annealing buffer [20 mM Tris–Cl (pH 7.6), 2 mM $MgCl_2$, 50 mM NaCl] were annealed by heating to 95°C for 2 min and then slowly cooling to room temperature. DNA was isolated by ethanol precipitation. When indicated, 5-hydroxymethylated cytosines within the annealed, ds DNA fragments were glucosylated by treatment with β-glucosyltransferase (New England Biolabs) according to the company protocol. After phenol extraction, the DNA was isolated by ethanol precipitation and dissolved in TE. DNA concentrations were determined from their absorbance at 260 nm. Electrophoresis of an aliquot on native 20% polyacrylamide gels followed by staining with ethidium bromide confirmed the determined concentrations.

For radiolabeled DNA, single-stranded oligomers were treated with T4 polynucleotide kinase (Optikinase; USB) in the presence of [γ-^{32}P]ATP. After annealing, the DNA was purified using Microspin G-25 chromatography (GE Healthcare Life Sciences).

The transcription template, pDKT90, which contains the T4 middle promoter P_{uvsX}, has been described (25).

Crystallographic analyses

The annealed duplexed DNA was diluted five-fold in buffer B and mixed with purified $MotA^{linker/CTD}$ at a 1.15:1 DNA:protein molar ratio. Prior to crystallization, the complex was concentrated to 10 mg/ml using Amicon 30 NMLW centrifugal concentrators (Millipore). Initial crystallization trials were performed with commercially available screening kits, and optimization generated the final crystals in 23% PEG 8000, 0.12 M sodium acetate, 0.1 M sodium cacodylate pH 6.5, 3% glycerol. Crystals were in space group $P6_1$ with cell parameters of $a = 72.27$ Å, $b = 72.27$ Å, $c = 279.37$ Å, $\alpha = 90.0°$, $\beta = 90.0°$, $\gamma = 120.0°$.

Crystals were cryoprotected by quick passage through mother liquor containing 25% glycerol and flash-frozen in liquid nitrogen. Diffraction data were collected at 100 K from a single crystal at SER-CAT beam line 22-ID (APS, Argonne National Laboratory) at 1.00 Å wavelength and processed to 2.96 Å using XDS (26) (Table 1). The coordinates for $MotA^{CTD}$ (1KAF) and an ideal B-form DNA double helix with the appropriate sequence generated using COOT (27) were used for molecular replacement using PHASER (28). This identified a solution with two copies of a $MotA^{linker/CTD}$–DNA complex in the asymmetric unit. However, refinement using PHENIX.REFINE (29) failed to improve the model and large holes in the structure were inconsistent with a stably packed crystal lattice. Weak electron density evident from composite omit maps was eventually interpreted as two additional copies of $MotA^{linker/CTD}$ in the crystal asymmetric unit and the structure was successfully refined using iterative cycles of PHENIX.REFINE and rebuilding using COOT. 5% of the reflections were excluded from the refinement process for the calculation of Rfree. Diffraction anisotropy (twin law K,H,-L) was identified using the 'Diffraction Anisotropy Server' (http://services.mbi.ucla.edu/anisoscale) (30), and this was automatically corrected during the refinement with PHENIX.REFINE. The ends of the two DNA molecules in the asymmetric unit were very clear in the electron density map; there is an unambiguous discontinuity in the electron density and a ∼20° rotational shift between the ends of the adjacent double helices (Supplementary Figure S1). However, the orientation of each duplex at this resolution was less obvious. To verify the modeled DNA orientation, 10 cycles of refinement were performed on the final model with discriminatory base pairs G9-C14′, C10-G13′ and T11-A12′ removed from both DNA duplexes. $F_o - F_c$ omit maps confirmed that both duplexes bind the two copies of $MotA^{linker/CTD}$ in the same orientation (Supplementary Figure S2). The final model comprised two DNA helices, four copies of $MotA^{linker\text{-}CTD}$ and 389 water molecules with R_{work} and R_{free} values of 22.0% and 24.5%, respectively (Table 1).

Docking of α- and β-D-glucosyl-5-hydroxymethyl-deoxycytosine

The coordinates of α-D-glucose and β-D-glucose bound to a hydroxymethyl-deoxycytosine were generated using AVOGADRO/BABEL (31). The corresponding restraints were generated using PHENIX/ELBOW (32) in order to manually place the grafted cytosine with COOT in the identified position in the crystal structure of the $MotA^{linker/CTD}$-DNA complex. Optimization and energy minimization of the structure was performed us-

Table 1. Data collection and refinement statistics

Data collection	
Space group	$P6_1$
Cell dimensions	
a, *b*, *c* (Å)	72.27, 72.27, 279.37
α, β, γ (°)	90, 90, 120
Resolution (Å)	93.12–2.96 (3.11–2.96)[a]
R_{merge}	0.167 (1.883)[b]
R_{meas}	0.171 (1.924)[b]
R_{pim}	0.036 (0.376)
I/σ (*I*)	15.1 (2.0)
$CC_{1/2}$	0.999 (0.828)
Completeness (%)	100.0 (100.0)
Redundancy	23.4 (23.5)
Anisotropy score (%)[c]	31.2
Refinement	
Resolution (Å)	62.59–2.96
No. reflections	17 252
R_{work}/R_{free}	22.0/24.5 (33.6/34.8)
Ramachandran plot	
Outliers (%)	1.6
Allowed (%)	2.3
Favored (%)	96.1
Clashscore	18.9
Molprobity score	2.05
Twinning fraction (%)	28
No. atoms	
Protein and DNA	5350
Water	389
B factors	
Protein and DNA	90.9
Water	67.6
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.544

[a]Values in parentheses are for the highest-resolution shell.
[b]These values are high in the outer shell because of the very high data redundancy.
[c]Obtained from the 'diffraction anisotropy server' (http://services.mbi.ucla.edu/anisoscale).

ing SCHRODINGER/MAESTRO (Maestro, Schrödinger, LLC, New York, NY, 2017).

In vitro transcription

Transcription reactions (5 μl) contained 0.25 pmol σ^{70}, 0.05 pmol core RNAP, 1.7 pmol AsiA, 1.8 pmol MotA, 0.01 pmol DNA (pDKT90 linearized with BsaAI), 1× Kglu transcription buffer [40 mM Tris–acetate (pH 7.9), 150 mM sodium glutamate, 4 mM magnesium acetate, 0.1 mM EDTA, 0.1 mM DTT, 100 μg/ml BSA], 200 μM ATP, 200 μM GTP, 200 μM CTP, 5 μM [γ-^{32}P] UTP (~1 × 10^5 dpi/pmol), and 500 ng heparin. AsiA and σ^{70} were first incubated at 37°C for 10 min and then incubated with core RNAP for an additional 10 min at 37°C. Transcription reactions were then assembled on ice by adding the AsiA-associated RNAP to MotA and DNA. After 1 min at 37°C, the ribonucleoside triphosphates and heparin were added to initiate a single round of transcription. Reactions were incubated for 8.5 min at 37°C, collected on dry ice, and processed as described (14).

Electrophoretic mobility shift assays (EMSA)

EMSA experiments were performed essentially as described (14) except that reactions contained 0.01 pmol of the 19 bp 5′-^{32}P labeled fragment (5′AATT**ATTTGCTTTA**GATT A3′, either unmodified or modified at the cytosine; MotA box motif indicated in bold), 1.7 pmol MotA (unless indicated otherwise), and as indicated, unlabeled competitor DNA. [Note that this MotA box sequence (5′ATTTG CTTTA) differs from that used for crystallography (5′C TTTGCTTAA); it was chosen because extensive EMSA analyses have been previously performed using this DNA (33).] Reactions were subjected to electrophoresis in native, 12% acrylamide gels run in 1× TBE at 4°C for 2.5–3 h at a constant voltage of 150 V. After autoradiography, films were scanned using a Powerlook 2100XL densitometer, and quantification was performed using Quantity One software from Bio-Rad. For competition experiments, values represent (% bound in the presence of competitor)/(% bound in the absence of competitor). $K_{d(app)}$ values were determined from the amount of protein needed to bind 50% of the DNA (33).

Surface plasmon resonance (SPR) assays

The general procedures used were the same as those previously described (21). The 22 bp 5′-^{32}P labeled fragment 5′GAAG**CTTTGCTTAA**TAATCCAC3′ (MotA box motif indicated in bold) was either unmodified or modified by the addition of a glucosylated, hydroxymethyl moiety on one or both cytosines within the MotA box motif. The top strand of each double-strand oligomer also contained the single-stranded sequence 5′ biotin-ATACATTATG to allow attachment of the DNA to the neutravidin present on the SPR chip.

RESULTS

Structure determination of the MotA$^{linker/CTD}$–DNA complex

Using surface plasmon resonance (SPR), we have previously identified a 22-mer double-stranded (ds) DNA segment containing the MotA box motif that binds full-length MotA with a K_D(app) of 137 nM (21). Previous work has indicated that the MotA^{CTD} can interact with MotA box DNA (23), and our recent modeling suggested that the MotA linker also contributes to binding. Consequently, with the goal of optimizing the MotA construct for crystallization of the complex, we repeated the SPR experiment using the same dsDNA segment and a MotA$^{linker/CTD}$ construct encompassing residues 93–211, which includes residues 93–104 from the linker. This showed that the MotA$^{linker/CTD}$ construct binds with a slightly improved K_D of 93 nM (Supplementary Figure S3).

We successfully crystallized the more stable complex and the structure was determined by molecular replacement (MR) at 2.96 Å resolution. The final model had R_{work}/R_{free} values of 22.0/24.5 and acceptable geometry for this resolution (Table 1). The crystal asymmetric unit comprised two identical copies of a 2× MotA$^{linker/CTD}$–DNA complex (Figure 1A). Within each complex, one

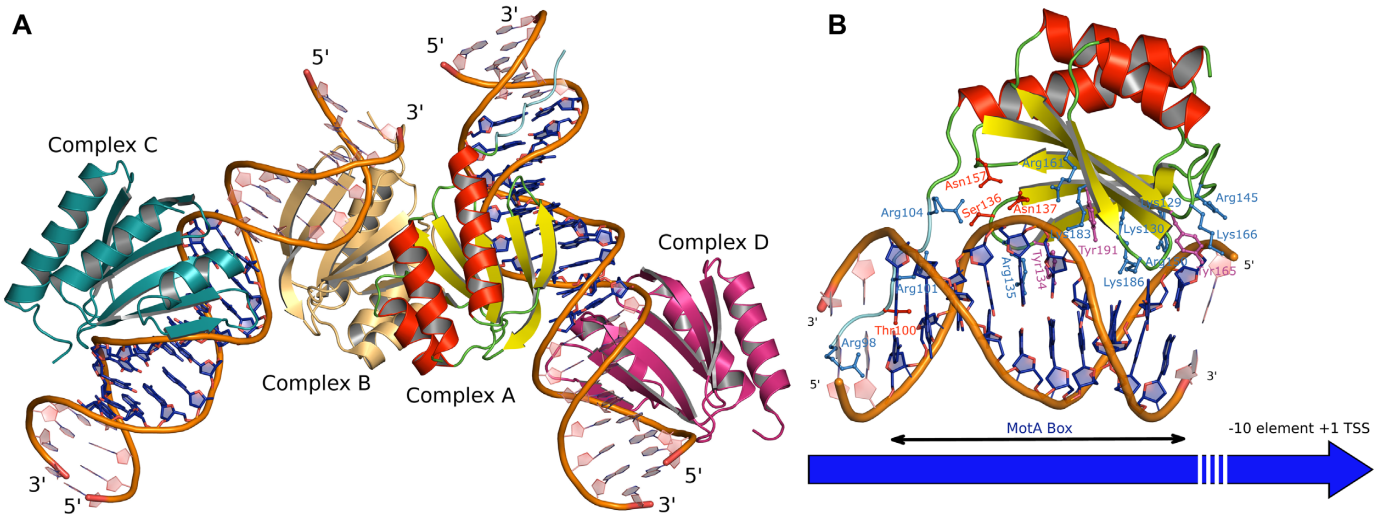

Figure 1. The MotA$^{linker/CTD}$–DNA assembly. (**A**) The complete assembly in the structure comprises two cognate complexes, complex A (yellow strands, red helices and green loops) and complex C (light blue), and two non-cognate complexes, complex B (gold) and complex D (magenta). (**B**) The isolated complex A. The blue arrow indicates the direction of transcription relative to the complex, and the locations of the –10 element and the Transcription Start Site (TSS). The key interacting residues are; blue – basic, red – polar, magenta – aromatic. A schematic representation of complex A is shown in Supplementary Figure S5**.** In both figures, the linker region in complex A is shown in light blue and the base pairs corresponding to the MotA box motif are shown in dark blue.

MotA$^{linker/CTD}$ is bound specifically at the MotA box motif and the second is bound separately and non-specifically 180° away on the opposite side of the dsDNA. A cognate MotA$^{linker/CTD}$ from one complex forms a dimer with the non-cognate MotA$^{linker/CTD}$ from the adjacent complex to create the crystal asymmetric unit. As observed in the MotA^{CTD} crystal structure without DNA (20), the double-wing domain has a propensity to form side-by-side multimers via the six-stranded β-sheet, which explains this packing scheme. In this assembly, the specific MotA$^{linker/CTD}$–DNA complex that mediates the non-crystallographic dimer will be referred to as complex A and the non-specific MotA$^{linker/CTD}$/DNA complex on the same DNA is complex B. Complexes C and D are the corresponding specific and non-specific complexes, respectively, on the other DNA in the asymmetric unit (Figure 1A).

Description of the MotA$^{linker/CTD}$–DNA complex

Of the two cognate structures, complex A is better resolved and more complete than complex C and has lower overall B factors (Supplementary Figure S4). Complex A will therefore form the basis of the structural description (Figure 1B, Supplementary Figure S5). The central and conserved 5′-GCT-3′ sequence within the MotA box motif is in the 'top strand' of the DNA construct at positions 9, 10 and 11, and their 5′-A'G'C'-3′ partners on the 'bottom strand' are at positions 12′, 13′ and 14′. The overall structure of MotA^{CTD}, which contains three α-helices and a 6-stranded anti-parallel β-sheet, is almost identical to that previously described (20) with the only changes limited to the surface side chain orientations. The domain engages the DNA across the major groove via the β-sheet such that the outer strand β1 spans the sugar-phosphate backbones while the opposite outer strand β5 sits along the sugar-phosphate backbone of the top strand, N to C, 5′ to 3′. The binding of MotA$^{linker/CTD}$ is centered on the MotA box motif and is mediated mostly by electrostatic and hydrogen-bonding interactions with the phosphate groups. Ser136, Asn137, Asn157, Arg161, Lys183, Tyr191 and potentially Lys186 interact with the top strand (Figure 2A, Supplementary Figure S5), and Lys129, Arg145, Arg150, Tyr165, Lys166 and potentially Lys130 interact with the bottom strand (Figure 2B, Supplementary Figure S5).

The MotA box motif is only formally 'read' by three side chains, Tyr134 and Arg135 on the loop between strands β1 and β2, and Arg101 in the N-terminal linker. The hydroxyl group of Tyr134 forms a hydrogen bond with the N4 nitrogen atom of cytosine C10 in the C10-G13′ base pair, and the adjacent Arg135 is hydrogen bonded to the guanine G9 of the G9-C14′ base pair (Figure 2C and Supplementary Table S2). Arg101 is hydrogen bonded to the O2 oxygen atom of thymine T6 in the T6-A17′ base pair in the minor groove, which emphasizes the key role of the linker in the cognate complex (Figure 2D). Consistent with our previous studies (21,33), the linker region has a number of additional interactions with the minor groove (Figure 2D). Thr100, Lys102 and Arg104 interact with phosphate groups, and Arg98 appears to penetrate the minor groove to interact with base pairs A3-T20′/G4-C19′ although the electron density at this peripheral region is quite weak. The linker region in complex C is less well resolved but the visible proximal region also engages the minor groove with Arg104 clearly interacting with adjacent phosphate groups.

The two non-specific complexes B and D are very similar to the two specific complexes. This is not surprising because they are mediated by the same side interactions with the DNA phosphate groups. This similarity is evident when all four complexes are superimposed (Supplementary Figure S6). Closer examination reveals that both complexes are partially specific because there is a fortuitous GC base pair (G4′-C19) in each that interacts with Arg135 in the same

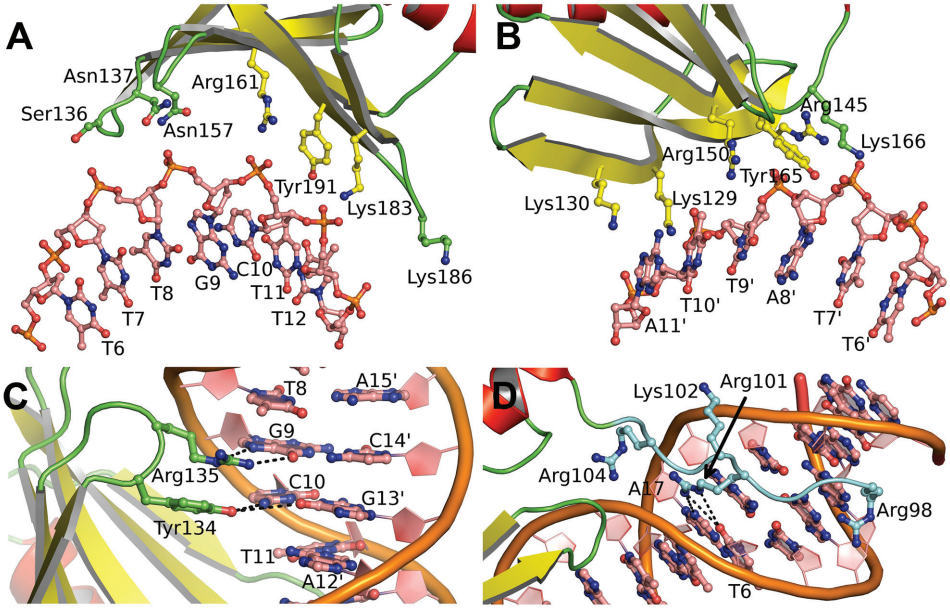

Figure 2. Structural details of the specific (cognate) MotA$^{linker/CTD}$–DNA complex. The complex shown is complex A, although the interactions in complex C are almost identical. The side chain interactions with the 'top' DNA strand (**A**) and the 'bottom' DNA strand (**B**). (**C**) Recognition of the MotA box motif by Tyr134 and Arg135. (**D**) Specific and non-specific interactions of the linker region (light blue or cyan) with the minor groove upstream of the MotA box motif. Note the interaction of Arg101 with T6-A17′ in the MotA box motif (shown with an arrow).

fashion that this residue interacts with the G9-C14′ base pair in the MotA box motif. However, Tyr134 is opposite an AT base pair and makes no interaction with the bases (Supplementary Table S2). Complex B is less tight than the other three, which is probably a result of it being less well packed in the crystal lattice. The role, if any, of the linker region in the non-specific complexes cannot be distinguished in the structure because the DNA construct does not extend far enough to permit this interaction.

A polar cavity accommodates 5-glucosylated, hydroxymethylated cytosine

Wild type T4 DNA is modified by both α- and β-D-glucosyltransferases, which add a glucosylated, hydroxymethyl group at position 5 of cytosines. This modification results in a sugar moiety within the major groove of the DNA, which protects the phage genome from host nucleases as well as T4-encoded nucleases that target host DNA (34). In our DNA construct, C10 within the MotA box motif is directly at the protein-DNA interface, and a distinct electropositive cavity adjacent to the carbon 5 is ideally positioned to accommodate the glucosylated, hydroxymethyl moiety (Supplementary Figure S7A). To test this proposal, the coordinates for the moiety were modeled onto the five carbon of C10 to generate either α- or β−D-glucosyl-5-hydroxymethyl-deoxycytosine, and then computationally optimized and energy minimized in complex A of the crystal structure. For both modifications, the model predicts that the cavity accommodates the sugar group by forming hydrogen bonding interactions with Glu153 and P9 of the DNA and packing against the side chains of Tyr134 and Arg135 that recognize the adjacent MotA box motif (Figure 3A, B and Supplementary Figure S7B).

This model suggested that the presence of the modification on the top strand of the DNA should provide additional stability for the binding of MotA, while a modification on the bottom strand should have no effect because it does not contact the bound protein. To directly test this, we performed electrophoretic mobility shift assays (EMSAs) to determine the ability of various modified DNA fragments to compete with a 5′-^{32}P end-labeled, 19 base pair fragment containing unmodified MotA box DNA. We only used the β-D-glucosyl-5-hydroxymethyl-deoxycytosine modification for these experiments because the α-D-glucosyltransferase enzyme is unavailable. We first tested the following fragments: no modification, hydroxymethyl moieties on both Cs within the MotA box motif, and glucosylated, hydroxymethyl modification on both C's (Figure 4A). Fully modified DNA was significantly more competitive than either the unmodified or the hydroxymethyl modified fragments. As expected (35), a mutant fragment that does not contain a MotA box motif did not compete. To investigate the effect of top versus bottom strand modification, we then generated ds oligomers with top or bottom strand modifications, either directly (Figure 4B and Figure 4C) or by re-annealing fully modified ds fragments in the presence of a 10-fold excess of ssDNA that was unmodified on the top or bottom strand (Supplementary Figure S8A). These EMSA's indicated that the primary effect on MotA binding was due to modification on the top strand as predicted. Finally, we used EMSA to obtain $K_{D(app)}$ values for the unmodified fragment (∼140 nM) and the top strand modified fragment (∼20 nM) (Supplementary Figure S8B). However, the binding curve observed with the modified DNA was quite steep, making it difficult to reliably determine the $K_{D(app)}$ from this analysis.

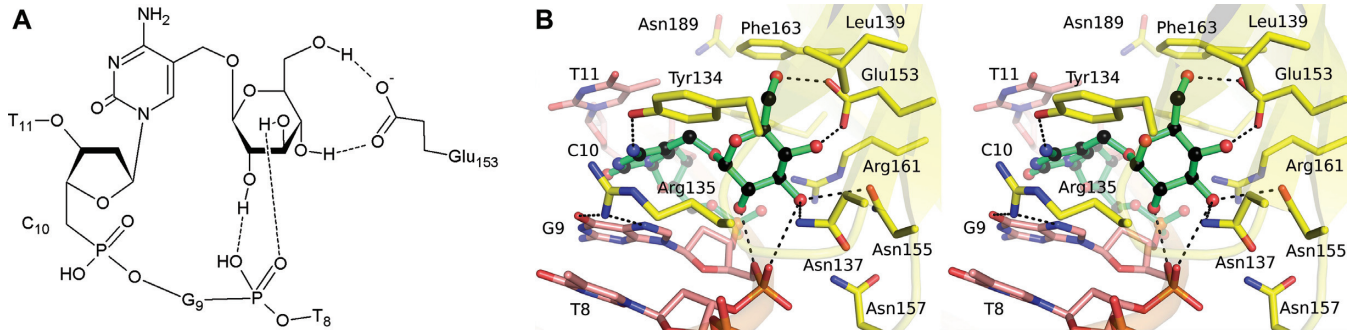

Figure 3. A buried cavity in the MotA$^{linker/CTD}$ –DNA specific complex is suitably located to accommodate the sugar moiety on the modified cytosine within the MotA box motif. (**A**) Schematic representation of the docked β-D-glucosyl-5-hydroxymethyl-deoxycytosine into the protein cavity showing the predicted hydrogen bonding network with the surrounding residues. (**B**) Stereo view of the actual docked structure; side chain carbon atoms - yellow stick; DNA carbons – rose stick; β-D-glucosyl-5-hydroxymethyl cytosine - ball-and-stick with black carbons and green bonds. The structure of the docked α-D-glucosyl-5-hydroxymethyl cytosine is very similar and shown in Supplementary Figure S7B.

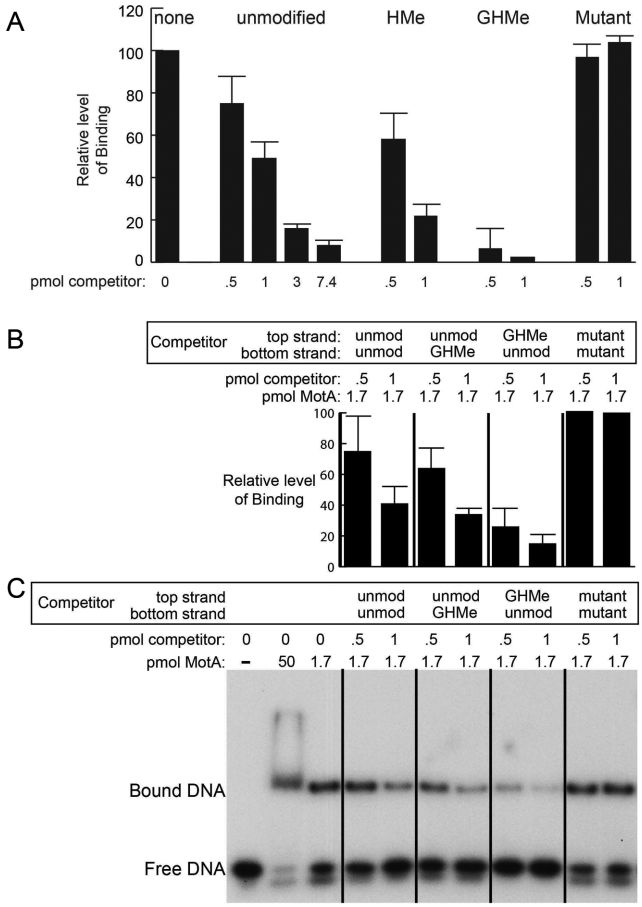

Figure 4. Electrophoretic mobility shift assays (EMSAs) to analyze the MotA$^{linker/CTD}$ –DNA binding interaction. (**A**) MotA box DNA containing glucosylated, hydroxymethylated cytosines (GHMe) competes better than unmodified or hydroxymethyl modified (HMe) MotA box DNA for MotA binding. A mutant fragment that was previously found to not bind MotA (5′AATTATTATACTTAGATTA3′) (35) does not compete. Values and standard deviations were obtained from three to seven experiments. (**B**) Modification of the top strand cytosine within the MotA box by GHMe significantly improves MotA binding. Values and standard deviations were obtained from three to six experiments except for competitions with the mutant competitor, which were performed twice. (**C**) Representative native acrylamide gel from (**B**).

To further confirm that the glucosyl-hydroxymethyl modification stabilizes the specific MotA-DNA interaction, we used SPR to directly measure the binding. We used the same general procedure as described in our earlier studies (21). Unmodified DNA bound to full-length MotA with a K_D of 117 nm, consistent with the value determined by EMSA (Figure 5A), and DNA with the modification on the bottom strand only showed a modestly improved K_D of 66 nm (Figure 5B). However, DNA modified on both strands (Figure 5C) or only on the top strand (Figure 5D) displayed drastically improved binding with K_D values of 0.285 nM and 0.359 nM, respectively. These SPR data indicated that the K_D with modified DNA is in fact even lower than was determined by EMSA. The SPR data also revealed that the increased affinity is due to much slower dissociation rates when the appropriately placed modification is present in the complex (Supplementary Table S3).

Mutagenesis studies on the MotA$^{linker/CTD}$–DNA complex

To evaluate the importance of the residues in the cavity that we propose interact with base determinants (Tyr134 with C10 and Arg135 with G9) or with the glucosyl-hydroxymethyl modification (Glu153), we constructed and purified proteins with the single point substitutions Y134A, R135A and E153A, as well as the double mutant Y134A/R135A. We then performed *in vitro* transcription analyses, EMSAs and SPR binding studies. The transcription analyses were performed on unmodified DNA templates and the single mutations had minimal effect on activity (Supplementary Figure S9A). This result is not unexpected because the complete transcription complex involves multiple interacting components and a reduced MotA–DNA interaction would be predicted to have a minimal effect. We note that previous work has identified active MotA-dependent promoters that deviate from consensus at both the G9 and C10 positions, which would also be expected to significantly reduce binding affinity (35,36). The 2-fold drop in activity for Y134A/R135A is significant and suggests weaker binding to the MotA-box motif. However, it may also reflect that the loss of two central side chains impacts the folded structure of the protein.

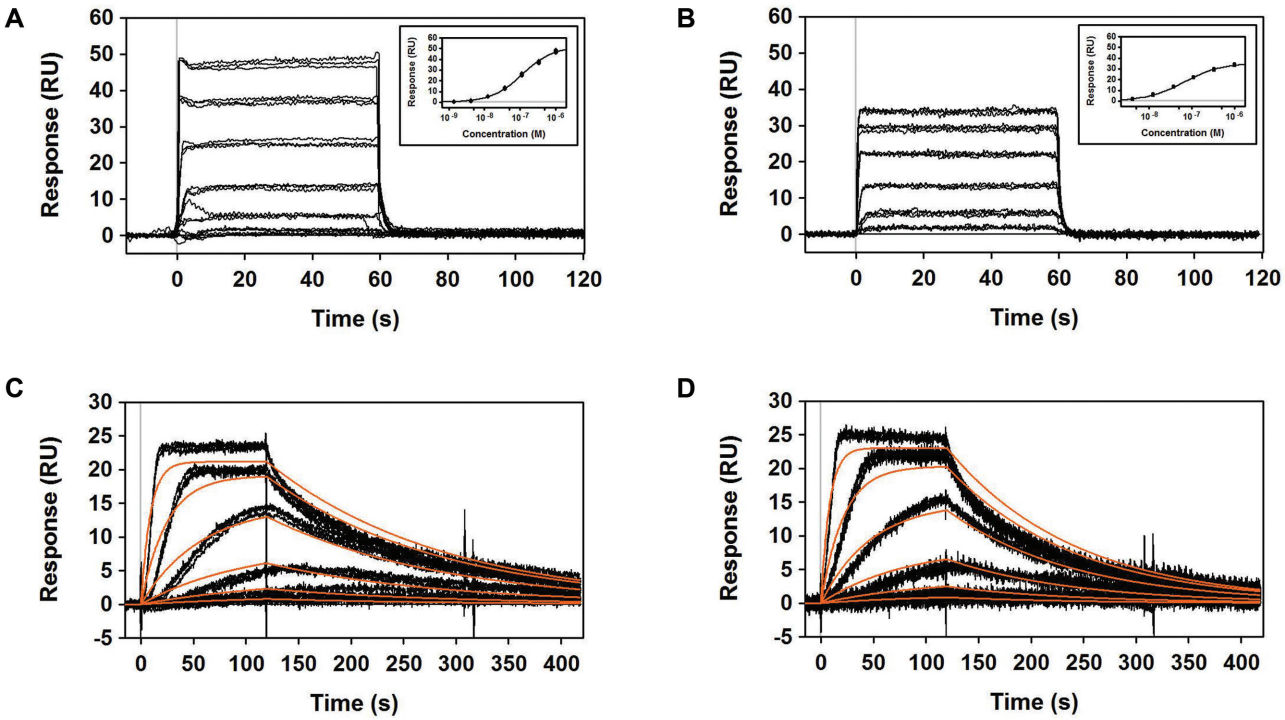

Figure 5. Surface plasmon resonance analyses of the MotA$^{linker/CTD}$–DNA binding interaction. The 'top' strand of the immobilized dsDNA contained the sequence 5′-GAAG**CTTTGCTTAA**TAATCCAC-3′ (the MotA box motif is shown in bold). (**A**) MotA + unmodified/unmodified DNA (neither DNA strand is modified). (**B**) MotA + unmodified/GHMe DNA (only the bottom DNA strand is modified). (**C**) MotA + GHMe/GHMe DNA (both DNA strands are modified). (**D**) MotA + GHMe/unmodified DNA (only the top DNA strand is modified). In each panel, the response is shown for increasing concentrations of MotA$^{linker/CTD}$. The inset windows in (A) and (B) and the orange lines in (C) and (D) show the Langmuir fits to a 1:1 kinetic model. See Supplementary Table S3 for kinetic and affinity constants.

The EMSA (Supplementary Figure S9B) and SPR analyses (Supplementary Table S4) are far more sensitive indicators of the isolated MotA-DNA interaction. In EMSAs using unmodified DNA, we observed serious binding defects for the single R135A variant and the Y134A/R135A double mutant. This was confirmed by SPR that showed no detectable binding with either fully unmodified or unmodified DNA on the top strand and severely impaired binding for fully modified or DNA modified on the top strand. In addition, the EMSAs revealed that binding of the unmodified DNA was less stable with these variant proteins, as evidenced by the smear of radioactivity extending from the bound species to the free DNA (Supplementary Figure S9B). Once again, the Y134A/R135A impairment may be due to protein misfolding but Arg135 clearly has a key role in the interaction. Although Y134A protein behaved similarly to wt in the EMSA, SPR revealed impaired DNA binding, but to a lesser extent than R135A. Finally, the EMSA and SPR analyses of E153A, which we predict should selectively affect the interaction with the DNA modified on the top strand, were equivocal. The EMSAs showed a slightly tighter binding to DNA and no difference with DNA modified on the top strand (data not shown), and SPR also revealed a significantly tighter binding to unmodified DNA. In our model shown in Figure 3, Glu153 forms two hydrogen bonds with the glucose moiety of the DNA modifying group, but the van der Waals interaction within the large cavity is far more extensive. We suggest that the tighter binding of E153A with DNA results from the loss of the negative charge and increased electrostatic interaction with the DNA backbone. This is consistent with the observation (see Materials and Methods) that MotAE153A elutes from the phosphocellulose column at a higher salt concentration.

The DNA structure is minimally distorted in the MotA$^{linker/CTD}$–DNA Complex

Compared to standard B-form DNA, the DNA conformation in the MotA$^{linker/CTD}$ –DNA complex is largely undistorted by the interaction with the protein (Supplementary Figure S10). The DNA coordinates within the complex were uploaded to the W3DNA web server http://w3dna.rutgers.edu (37) to analyze the DNA conformation, and the key parameters are shown in Supplementary Table S5. The only distortions of any significance occur in the top strand where the MotA box motif interacts with Tyr134 and Arg135, the DNA backbone interacts with multiple side chains, and the linker region penetrates the minor groove. In the latter region, the DNA backbone moves towards the protein by ∼2.2 Å and the minor groove opens by ∼2.3 Å. Thus, MotA$^{linker/CTD}$ is ideally structured to recognize the B-DNA conformation, which explains how it can also form a non-specific complex, and is consistent with previous biochemical analyses indicating that binding by MotA does not significantly distort the DNA (33). We recognize that the DNA may become more distorted when the complex accommodates 5-glucosylated, hydroxymethylated cytosine,

but the open binding cavity for this moiety suggests that this would be minimal. It is also surprising that the DNA is not bent towards the bound protein because many of the negatively charged phosphate groups at the interface interact with neutralizing basic residues. However, this may be a crystal artifact because the second copy of $MotA^{linker/CTD}$ is bound to the opposite face of the DNA, which may cancel out any bending towards the protein in the opposite complex.

DISCUSSION

Our work has identified a new protein-DNA complex in which a 6-stranded anti-parallel β-sheet engages and straddles the major groove. The binding is mediated by two sets of basic and polar residues on either side of the β-sheet that interact with the two sugar-phosphate backbones. In contrast to modules like the helix-turn-helix in which an α-helix penetrates the major groove and 'reads' the base pairs, $MotA^{CTD}$ uses two residues from the surface of the β-sheet and one residue from the linker region to access the cognate MotA box motif via the major and minor grooves, respectively. Another feature of the complex is the use of a basic and flexible short polypeptide to augment the binding by interacting with the DNA minor groove using both electrostatic and hydrogen bonding interactions with the DNA backbone and interactions with the bases via arginine side chains. We previously tried unsuccessfully to crystallize the complex using $MotA^{CTD}$ constructs lacking the N-terminal linker, and the inclusion of this element prompted by our earlier studies (21) proved to be pivotal. The binding of arginine side chains into the minor groove has recently been shown to be an important shape/specificity determinant factor in protein-DNA complexes (38). It is typically associated with AT-rich tracks and a compressed minor groove, and although Arg101 does engage the T6-A17′ base pair within the 5′-T6-T7-T8-3′ sequence, the minor groove is not obviously compressed in this region.

Our crystal structure also shows that this mode of binding can create a reasonably tight non-specific complex. Our DNA construct is longer than is required for the specific $MotA^{linker/CTD}$ complex, but this allowed a second $MotA^{linker/CTD}$ to bind independently on the opposite face of the DNA. Thus, it was possible to visualize both cognate and non-cognate complexes in our structure. This is not altogether surprising given the previous work indicating that MotA accepts many deviations from its consensus sequence (35,36,39,40). In addition, although the MotA box motif consensus sequence contains a highly conserved GC base pair within its center (9), previous work has indicated that unmodified DNA is fully competent for MotA binding (25) and MotA/AsiA activation of transcription (24,41). Furthermore, MotA does not require a specific base determinant at the GC base pair and can accept base substitutions at these positions (33).

Although AsiA and MotA analogs have been identified in dozens of T4-like phages, a process similar to σ appropriation has not been observed outside phage. However, the novel double-wing structure has been reported for two proteins of unknown function, the conserved *E. coli* YjbR (42) and the *Pseudomonas syringae* protein Pspto_3016 (20,43). Thus, an understanding of how MotA interacts with its partner DNA will yield important insights into the functions of these proteins and other as yet unidentified members of the family. Another protein that contains this domain is the phage T4 SF2 helicase UvsW. UvsW has a key role in T4 recombination and is functionally and structurally related to the eukaryotic Rad54 protein (44,45). It also contains a double-wing subdomain and the $MotA^{linker/CTD}$ complex now provides insights into its role. The largely non-specific interactions via the sugar-phosphate backbone in the complex may allow MotA to search for the MotA box motif by 'riding along' the DNA via the saddle-like surface of the double-wing domain (Supplementary Figure S11A). UvsW translocates and remodels three-way and four-way DNA junctions, and the MotA-like domain is ideally positioned to bind one of the translocating dsDNA arms (Supplementary Figure S11B). The HARP domain of the SF2 helicase SMARCAL1 is structurally similar to the MotA-like domain of UvsW and has been shown to have a similar function (46). Other examples of protein–DNA interactions that are mediated by β-sheets include the TATA-binding protein (47), EcoRV (48), MutS (49,50) and the Met repressor (51).

The isolated $MotA^{linker/CTD}$–DNA complex observed in our crystal must be consistent with the structure for MotA/DNA within the entire MotA/RNAP/AsiA/DNA transcription complex. Structural models for the MotA/DNA complex (21) and the complete transcription complex (41) have been developed, and the agreement is excellent (Supplementary Figure S12). This also confirms that no significant change in the structure of the complex occurs when the transcription complex is formed. Most significantly, the linker region between the N- and C-terminal halves of MotA, located within the minor groove of the upstream portion of the MotA box motif, is crucial for connecting the $MotA^{NTD}$ domain and its protein partner, the C-terminus of σ^{70}, with $MotA^{CTD}$ present on the DNA. Thus, through this mechanism, the phage converts the specificity of the host RNA polymerase from recognition of its typical -35 element motif in host promoter DNA to the MotA box sequence present in T4 middle promoters. The extra affinity afforded by the modification present on T4 DNA serves to ensure even greater specificity in this process.

Finally, the docking of $MotA^{CTD}$ to the major groove of DNA via a β-sheet rather than a penetrating α-helix creates space at the interface that the module exploits to create a cavity that recognizes and binds a glucosylated, hydroxymethyl modification to the T4 DNA cytosines. Consequently, as evidenced by the EMSA and SPR binding experiments, what is a low specificity complex from a base determinant perspective becomes much more specific from the presence of the sugar moiety on the top strand of the cytosine at the center of the MotA box motif. Many eukaryotic transcription factors are required to recognize and bind target sites on DNA that have been chemically modified by epigenetic processes, and recent work has revealed many transcription factors that preferentially target sequences that contain methylcytosine (52). Thus, our work has wider implications for the ways in which DNA modifications can ex-

tend the uniqueness of small DNA motifs to facilitate the specificity of protein-DNA interactions.

DATA AVAILABILITY

Structures are available in the PDB under ID 5JLT.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Darcie Miller, Eric Enemark and Stefan Gajewski for crystallographic advice and Kenneth Kreuzer for his insights and support. The diffraction data were collected at Southeast Regional Collaborative Access Team (SERCAT) beam line 22-ID at the Advanced Photon Source, Argonne National Laboratory, and we thank SERCAT staff for their assistance.

FUNDING

Work in the White laboratory was supported by National Institutes of Health (NIH) [GM066934]; Cancer Center core grant [CA21765]; American Lebanese Syrian Associated Charities (ALSAC); Work in the Hinton laboratory is supported by the Intramural Research Program of the National Institutes of Health, NIDDK; Use of the Advanced Photon Source is supported by the U.S. Department of Energy [W-31-109-Eng-38]; Supporting SERCAT institutions may be found at www.ser-cat.org/members.html. Funding for open access charge: St Jude Children's Research Hospital.
Conflict of interest statement. None declared.

REFERENCES

1. Decker,K.B. and Hinton,D.M. (2013) Transcription regulation at the core: similarities among bacterial, archaeal, and eukaryotic RNA polymerases. *Annu. Rev. Microbiol.*, **67**, 113–139.
2. Murakami,K.S. (2015) Structural biology of bacterial RNA polymerase. *Biomolecules*, **5**, 848–864.
3. Lee,D.J., Minchin,S.D. and Busby,S.J. (2012) Activating transcription in bacteria. *Annu. Rev. Microbiol.*, **66**, 125–152.
4. Murakami,K.S. (2013) X-ray crystal structure of Escherichia coli RNA polymerase sigma70 holoenzyme. *J. Biol. Chem.*, **288**, 9126–9134.
5. Saecker,R.M., Record,M.T. Jr and Dehaseth,P.L. (2011) Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J. Mol. Biol.*, **412**, 754–771.
6. Feklistov,A., Sharon,B.D., Darst,S.A. and Gross,C.A. (2014) Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu. Rev. Microbiol.*, **68**, 357–376.
7. Paget,M.S. (2015) Bacterial sigma factors and anti-sigma factors: structure, function and distribution. *Biomolecules*, **5**, 1245–1265.
8. Hook-Barnard,I.G. and Hinton,D.M. (2007) Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. *Gene Regul. Syst. Biol.*, **1**, 275–293.
9. Hinton,D.M. (2010) Transcriptional control in the prereplicative phase of T4 development. *Virology journal*, **7**, 289.
10. Hinton,D.M. and Vuthoori,S. (2000) Efficient inhibition of Escherichia coli RNA polymerase by the bacteriophage T4 AsiA protein requires that AsiA binds first to free sigma70. *J. Mol. Biol.*, **304**, 731–739.
11. Lambert,L.J., Wei,Y., Schirf,V., Demeler,B. and Werner,M.H. (2004) T4 AsiA blocks DNA recognition by remodeling sigma70 region 4. *EMBO J.*, **23**, 2952–2962.
12. Kuznedelov,K., Minakhin,L., Niedziela-Majka,A., Dove,S.L., Rogulja,D., Nickels,B.E., Hochschild,A., Heyduk,T. and Severinov,K. (2002) A role for interaction of the RNA polymerase flap domain with the sigma subunit in promoter recognition. *Science*, **295**, 855–857.
13. Finnin,M.S., Hoffman,D.W., Kreuzer,K.N., Porter,S.J., Schmidt,R.P. and White,S.W. (1993) The MotA protein from bacteriophage T4 contains two domains. Preliminary structural analysis by X-ray diffraction and nuclear magnetic resonance. *J. Mol. Biol.*, **232**, 301–304.
14. Bonocora,R.P., Caignan,G., Woodrell,C., Werner,M.H. and Hinton,D.M. (2008) A basic/hydrophobic cleft of the T4 activator MotA interacts with the C-terminus of E.coli sigma70 to activate middle gene transcription. *Mol. Microbiol.*, **69**, 331–343.
15. Cicero,M.P., Alexander,K.A. and Kreuzer,K.N. (1998) The MotA transcriptional activator of bacteriophage T4 binds to its specific DNA site as a monomer. *Biochemistry*, **37**, 4977–4984.
16. Pande,S., Makela,A., Dove,S.L., Nickels,B.E., Hochschild,A. and Hinton,D.M. (2002) The bacteriophage T4 transcription activator MotA interacts with the far-C-terminal region of the sigma70 subunit of Escherichia coli RNA polymerase. *J. Bacteriol.*, **184**, 3957–3964.
17. Finnin,M.S., Cicero,M.P., Davies,C., Porter,S.J., White,S.W. and Kreuzer,K.N. (1997) The activation domain of the MotA transcription factor from bacteriophage T4. *EMBO J.*, **16**, 1992–2003.
18. Li,N., Zhang,W., White,S.W. and Kriwacki,R.W. (2001) Solution structure of the transcriptional activation domain of the bacteriophage T4 protein, MotA. *Biochemistry*, **40**, 4293–4302.
19. Finnin,M.S., Hoffman,D.W. and White,S.W. (1994) The DNA-binding domain of the MotA transcription factor from bacteriophage T4 shows structural similarity to the TATA-binding protein. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 10972–10976.
20. Li,N., Sickmier,E.A., Zhang,R., Joachimiak,A. and White,S.W. (2002) The MotA transcription factor from bacteriophage T4 contains a novel DNA-binding domain: the 'double wing' motif. *Mol. Microbiol.*, **43**, 1079–1088.
21. Hsieh,M.L., James,T.D., Knipling,L., Waddell,M.B., White,S. and Hinton,D.M. (2013) Architecture of the bacteriophage T4 activator MotA/promoter DNA interaction during sigma appropriation. *J. Biol. Chem.*, **288**, 27607–27618.
22. Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, **185**, 60–89.
23. Gerber,J.S. and Hinton,D.M. (1996) An N-terminal mutation in the bacteriophage T4 motA gene yields a protein that binds DNA but is defective for activation of transcription. *J. Bacteriol.*, **178**, 6133–6139.
24. Hinton,D.M., March-Amegadzie,R., Gerber,J.S. and Sharma,M. (1996) Characterization of pre-transcription complexes made at a bacteriophage T4 middle promoter: involvement of the T4 MotA activator and the T4 AsiA protein, a sigma 70 binding protein, in the formation of the open complex. *J. Mol. Biol.*, **256**, 235–248.
25. March-Amegadzie,R. and Hinton,D.M. (1995) The bacteriophage T4 middle promoter PuvsX: analysis of regions important for binding of the T4 transcriptional activator MotA and for activation of transcription. *Mol. Microbiol.*, **15**, 649–660.
26. Kabsch,W. (2010) Xds. *Acta Crystallogr. D: Biol. Crystallogr.*, **66**, 125–132.
27. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
28. 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.
29. Afonine,P.V., Grosse-Kunstleve,R.W., Echols,N., Headd,J.J., Moriarty,N.W., Mustyakimov,M., Terwilliger,T.C., Urzhumtsev,A., Zwart,P.H. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 352–367.
30. Strong,M., Sawaya,M.R., Wang,S., Phillips,M., Cascio,D. and Eisenberg,D. (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 8060–8065.

31. Hanwell,M.D., Curtis,D.E., Lonie,D.C., Vandermeersch,T., Zurek,E. and Hutchison,G.R. (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics*, **4**, 17.
32. Moriarty,N.W., Grosse-Kunstleve,R.W. and Adams,P.D. (2009) electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol. Crystallogr.*, **65**, 1074–1080.
33. Sharma,M., Marshall,P. and Hinton,D.M. (1999) Binding of the bacteriophage T4 transcriptional activator, MotA, to T4 middle promoter DNA: evidence for both major and minor groove contacts. *J. Mol. Biol.*, **290**, 905–915.
34. Miller,E.S., Kutter,E., Mosig,G., Arisaka,F., Kunisawa,T. and Ruger,W. (2003) Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.: MMBR*, **67**, 86–156.
35. Marshall,P., Sharma,M. and Hinton,D.M. (1999) The bacteriophage T4 transcriptional activator MotA accepts various base-pair changes within its binding sequence. *J. Mol. Biol.*, **285**, 931–944.
36. Stoskiene,G., Truncaite,L., Zajanckauskaite,A. and Nivinskas,R. (2007) Middle promoters constitute the most abundant and diverse class of promoters in bacteriophage T4. *Mol. Microbiol.*, **64**, 421–434.
37. El Hassan,M.A. and Calladine,C.R. (1998) Two distinct modes of protein-induced bending in DNA. *J. Mol. Biol.*, **282**, 331–343.
38. 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.
39. Truncaite,L., Piesiniene,L., Kolesinskiene,G., Zajanckauskaite,A., Driukas,A., Klausa,V. and Nivinskas,R. (2003) Twelve new MotA-dependent middle promoters of bacteriophage T4: consensus sequence revised. *J. Mol. Biol.*, **327**, 335–346.
40. Truncaite,L., Zajanckauskaite,A. and Nivinskas,R. (2002) Identification of two middle promoters upstream DNA ligase gene 30 of bacteriophage T4. *J. Mol. Biol.*, **317**, 179–190.
41. James,T.D., Cardozo,T., Abell,L.E., Hsieh,M.L., Jenkins,L.M., Jha,S.S. and Hinton,D.M. (2016) Visualizing the phage T4 activated transcription complex of DNA and E. coli RNA polymerase. *Nucleic Acids Res.*, **44**, 7974–7988.
42. Singarapu,K.K., Liu,G., Xiao,R., Bertonati,C., Honig,B., Montelione,G.T. and Szyperski,T. (2007) NMR structure of protein yjbR from Escherichia coli reveals 'double-wing' DNA binding motif. *Proteins*, **67**, 501–504.
43. Feldmann,E.A., Seetharaman,J., Ramelot,T.A., Lew,S., Zhao,L., Hamilton,K., Ciccosanti,C., Xiao,R., Acton,T.B., Everett,J.K. *et al.* (2012) Solution NMR and X-ray crystal structures of Pseudomonas syringae Pspto_3016 from protein domain family PF04237 (DUF419) adopt a "double wing" DNA binding motif. *J. Struct. Funct. Genomics*, **13**, 155–162.
44. Kerr,I.D., Sivakolundu,S., Li,Z., Buchsbaum,J.C., Knox,L.A., Kriwacki,R. and White,S.W. (2007) Crystallographic and NMR analyses of UvsW and UvsW.1 from bacteriophage T4. *J. Biol. Chem.*, **282**, 34392–34400.
45. Sickmier,E.A., Kreuzer,K.N. and White,S.W. (2004) The crystal structure of the UvsW helicase from bacteriophage T4. *Structure*, **12**, 583–592.
46. Mason,A.C., Rambo,R.P., Greer,B., Pritchett,M., Tainer,J.A., Cortez,D. and Eichman,B.F. (2014) A structure-specific nucleic acid-binding domain conserved among DNA repair proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 7618–7623.
47. Kim,J.L., Nikolov,D.B. and Burley,S.K. (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, **365**, 520–527.
48. Kostrewa,D. and Winkler,F.K. (1995) Mg2+ binding to the active site of EcoRV endonuclease: a crystallographic study of complexes with substrate and product DNA at 2 A resolution. *Biochemistry*, **34**, 683–696.
49. Lamers,M.H., Perrakis,A., Enzlin,J.H., Winterwerp,H.H., de Wind,N. and Sixma,T.K. (2000) The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. *Nature*, **407**, 711–717.
50. Obmolova,G., Ban,C., Hsieh,P. and Yang,W. (2000) Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature*, **407**, 703–710.
51. Somers,W.S. and Phillips,S.E. (1992) Crystal structure of the met repressor-operator complex at 2.8 A resolution reveals DNA recognition by beta-strands. *Nature*, **359**, 387–393.
52. Yin,Y., Morgunova,E., Jolma,A., Kaasinen,E., Sahu,B., Khund-Sayeed,S., Das,P.K., Kivioja,T., Dave,K., Zhong,F. *et al.* (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, **356**, eaaj2239.