Structure
Article

Recognition of AT-Rich DNA Binding Sites by the MogR Repressor

Aimee Shen,[1,3] Darren E. Higgins,[1,*] and Daniel Panne[2,*]
[1]Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA
[2]European Molecular Biology Laboratory, 6 Rue Jules Horowitz, 38042 Grenoble, France
[3]Present address: Department of Pathology, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, CA 94305, USA
*Correspondence: dhiggins@hms.harvard.edu (D.E.H.), panne@embl.fr (D.P.)
DOI 10.1016/j.str.2009.02.018

SUMMARY

The MogR transcriptional repressor of the intracellular pathogen *Listeria monocytogenes* recognizes AT-rich binding sites in promoters of flagellar genes to downregulate flagellar gene expression during infection. We describe here the 1.8 Å resolution crystal structure of MogR bound to the recognition sequence 5′ ATTTTTTAAAAAAAT 3′ present within the *flaA* promoter region. Our structure shows that MogR binds as a dimer. Each half-site is recognized in the major groove by a helix-turn-helix motif and in the minor groove by a loop from the symmetry-related molecule, resulting in a "crossover" binding mode. This oversampling through minor groove interactions is important for specificity. The MogR binding site has structural features of A-tract DNA and is bent by ~52° away from the dimer. The structure explains how MogR achieves binding specificity in the AT-rich genome of *L. monocytogenes* and explains the evolutionary conservation of A-tract sequence elements within promoter regions of MogR-regulated flagellar genes.

INTRODUCTION

Flagella are surface structures that are required for bacterial motility, adhesion to host cells, invasion, and virulence (Macnab, 2004; Ramos et al., 2004; Van Houdt and Michiels, 2005). Given their key role in bacterial infections, flagella are also prone to being recognized by host pattern recognition receptors such as Toll-like receptor 5 (Andersen-Nissen et al., 2005; Feuillet et al., 2006; Hayashi et al., 2001). To evade detection by the host immune response, bacterial pathogens frequently downregulate production of flagella shortly after colonization (Andersen-Nissen et al., 2005; Ramos et al., 2004). In the Gram-positive, facultative intracellular pathogen *Listeria monocytogenes*, this downregulation occurs primarily at mammalian physiological temperature (37°C and above) and during replication within the eukaryotic cell cytosol (Gründling et al., 2004). Repression of flagellar gene expression is mediated by the motility gene repressor, MogR (Gründling et al., 2004). Binding of MogR to operator sequences located within flagellar gene promoter regions directly represses transcription of all flagellar motility genes in a nonhierarchal manner (Shen and Higgins, 2006). In the absence of MogR, all flagellar motility genes are constitutively expressed at high levels in a temperature-independent manner. The resulting overexpression of flagellin (FlaA) by MogR-negative bacteria induces a chaining phenotype. This abnormal cellular physiology compromises the ability of *L. monocytogenes* to invade host cells (Shen and Higgins, 2006). Thus, MogR is a key mediator of the transition of *L. monocytogenes* from a ubiquitous extracellular bacterium to an intracellular pathogen.

MogR binding sites were originally identified by examining MogR-regulated flagellar gene promoter regions for conserved sequence motifs (Shen and Higgins, 2006). This analysis revealed that the palindromic sequence 5′ TTTTNNNNNAAAA 3′ occurred repeatedly in flagellar gene promoter regions. The intervening nucleotides (N) typically contained A:T base pairs. For example, the best characterized MogR binding site within the *flaA* promoter region contains the sequence 5′ TTTTTTAAA AAAA 3′. The sequence requirements of this binding site were confirmed using gel mobility shift analysis with wild-type and mutant *flaA* promoter region sequences (Shen and Higgins, 2006). As predicted from the sequence conservation, the flanking thymines and adenines (underlined) were essential for MogR binding and the intervening sequences were more permissive to nucleotide changes.

The 5′ TTTTNNNNNAAAA 3′ sequence occurs ~5000 times (one mismatch allowed; ~500 times with no mismatch allowed) within the upstream regions of genes in the AT-rich genome of *L. monocytogenes* (60% AT) (Shen and Higgins, 2006). Although previous microarray analyses revealed that the primary targets of MogR repression are flagellar gene promoters, how MogR discriminates its recognition sequences within flagellar promoter regions over the multitude of possible genomic binding sites remained unknown. Promoter regions of MogR-repressed flagellar genes frequently contain multiple MogR binding sites spaced 1, 2, or 3 helical turns apart, and previous analyses have shown that a minimum of two MogR binding sequences are required to repress *flaA* transcription in *L. monocytogenes* (Shen and Higgins, 2006). Therefore, it has been proposed that MogR requires a minimum of two 5′ TTTTNNNNNAAAA 3′ sequences separated by integral helical turns to function as a transcriptional repressor (Shen and Higgins, 2006).

In this paper, we investigated the molecular mechanism by which MogR recognizes its A-tract DNA binding sequence and thus functions as a master regulator of flagellar gene expression. A-tract binding sites are of particular interest because they are known to exhibit intrinsic DNA curvature, which can play

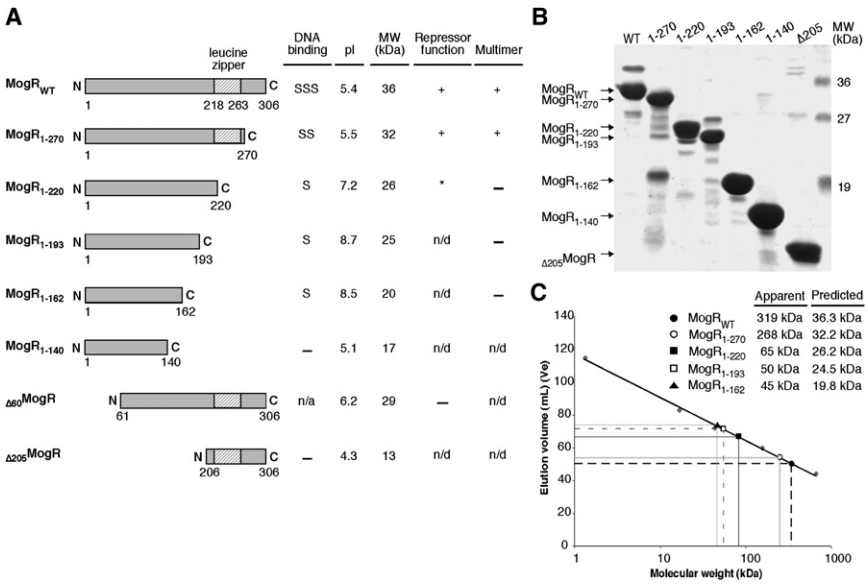

Figure 1. Analysis of MogR Deletion Constructs

(A) Schematic and summary of His_6-tagged MogR mutant proteins. The length of MogR remaining in the C-terminally truncated MogR proteins is indicated. For N-terminally truncated MogR proteins, the number of residues deleted from the N terminus is indicated following the Δ symbol. Gel mobility shift analysis was used to assess MogR binding to *flaA* promoter region DNA (Figure 2). The presence of super-supershifted (SSS), supershifted (SS), and shifted (S) DNA complexes is indicated. Note that the probe comprises the region −162 to +8 relative to the *flaA* transcriptional start site and contains multiple MogR binding sites. Repressor function was assessed by heterologously expressing *mogR* deletion constructs in MogR-negative bacteria from an ectopic locus and determining the ability of mutant MogR proteins to repress FlaA production (Figure S2). The asterisk denotes not determinable repressor function due to the destabilization of $MogR_{1-220}$ in *L. monocytogenes*. pI and MW indicate the calculated isoelectric point and molecular weight of MogR mutant proteins, respectively. MWs correspond to the predicted mass of His_6-tagged MogR proteins. Oligomerization (Multimer) was assessed by gel-filtration analyses (C). n/d indicates not done; n/a indicates not applicable due to the insolubility of $_{\Delta60}$MogR in *E. coli*.

(B) Coomassie stain of His_6-tagged MogR deletion mutant proteins. Equivalent volumes of eluate resulting from affinity purification of *E. coli* lysates containing His_6-tagged MogR variants were loaded.

(C) Gel filtration analysis of select MogR mutant proteins. The calibration curve based on molecular weight standards (gray diamonds) and the elution volumes of $MogR_{wt}$ (solid circle), $MogR_{1-270}$ (open circle), $MogR_{1-220}$ (solid square), $MogR_{1-193}$ (open square), and $MogR_{1-162}$ (solid triangle) have been plotted. The apparent and predicted molecular weights of the His_6-tagged MogR variants are given.

significant roles in transcriptional regulation by affecting promoter geometry (Crothers and Shakked, 1999). First, we determined that MogR contains two functional domains: an N-terminal DNA binding domain and a putative C-terminal leucine zipper oligomerization domain. In the absence of the C-terminal leucine zipper domain, MogR was monomeric and able to bind to its recognition site in vitro, albeit with reduced binding affinity compared to the full-length protein. We then determined the 1.8 Å resolution crystal structure of the N-terminal DNA binding domain of MogR bound to the *flaA* promoter region binding site 5′ ATTTTTTAAAAAAAT 3′. In the crystal structure, MogR is bound to its palindromic DNA target sequence as a dimer. The structure reveals that the whole 15 bp recognition element, except the central A:T base pair, is recognized and forms a composite MogR dimer binding site. These analyses provide insight into the mechanism of how an A-tract binding protein specifically recognizes its DNA binding sequence in the context of an AT-rich genome.

RESULTS

Identification of the MogR N-Terminal DNA Binding Domain

MogR was originally identified by virtue of its ability to bind to *flaA* promoter region DNA (Gründling et al., 2004). MogR possesses no apparent homology to any previously characterized protein, and sequence analysis did not reveal a DNA binding domain within MogR. The only apparent feature is a putative leucine zipper domain in the C terminus (amino acids 218–263; Figure 1A). To further identify functional domains within MogR, we designed C-terminal truncation constructs and assessed their ability to bind to *flaA* promoter region DNA using gel mobility shift assays and DNase I footprinting. These experiments showed that a truncated MogR protein containing residues 1–162 allowed DNA binding and that the C-terminal leucine zipper domain is important for overall binding affinity (Figure 2; see Figure S1 available online). The oligomerization status of these constructs was further investigated using size exclusion chromatography. We observed that MogR constructs containing the C-terminal leucine zipper region had significantly higher elution volumes, supporting the model that this domain is involved in MogR oligomerization (Figure 1C). The presence of a C-terminal oligomerization domain likely also explains the enhanced DNA binding affinity of constructs containing this domain (Figure 2; Figure S1).

To determine whether the functional domains of MogR identified in vitro play similar roles in *L. monocytogenes*, we measured the ability of MogR truncation constructs to repress FlaA protein production in MogR-negative *L. monocytogenes* (Δ*mogR*). Deletion of the N-terminal 60 amino acids abrogated FlaA repression (Figure S2, lanes 11–14). In contrast, deletion of the C-terminal 36 amino acids ($MogR_{1-270}$) still permitted FlaA repression, even though $MogR_{1-270}$ was produced at dramatically lower levels than wild-type MogR and also at lower levels than N-terminally truncated MogR (Figure S2, lanes 7–10). Further truncation of the C terminus ($MogR_{1-220}$) to remove the leucine zipper region destabilized the protein upon expression in *L. monocytogenes*, as $MogR_{1-220}$ could not be detected by western blot analysis. These findings are consistent with our in vitro analysis that localizes the DNA binding activity to the N terminus of MogR (Figure 2).

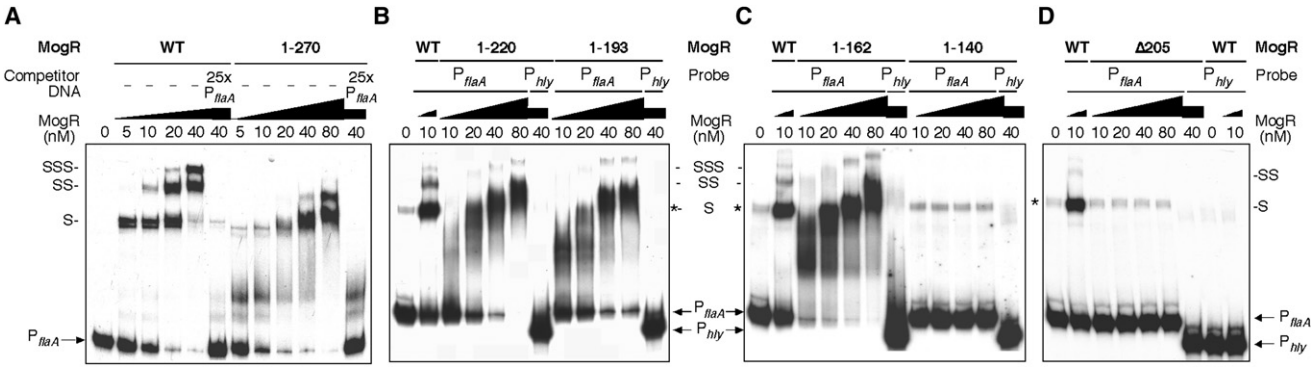

Figure 2. The N-terminal 162 Amino Acids of MogR Contain a DNA Binding Domain

Binding of His_6-tagged MogR mutant proteins to *flaA* promoter region DNA by gel mobility shift analyses. Radiolabeled *flaA* promoter region DNA was incubated with increasing concentrations of His_6-tagged MogR variants. The binding reactions were analyzed by nondenaturing PAGE. The positions of free probe, shifted (S), supershifted (SS), and super-supershifted (SSS) DNA complexes are indicated.

(A) $MogR_{wt}$ and $MogR_{1-270}$. Unlabeled competitor *flaA* probe DNA was added in 25-fold excess to the binding reactions where indicated.

(B–D) Binding of $MogR_{1-220}$, $MogR_{1-193}$, $MogR_{1-162}$, $MogR_{1-140}$, and $_{\Delta 205}$MogR mutant proteins to *flaA* and *hly* promoter region DNA was also assessed by gel mobility shift analysis. Asterisks (*) denote a radiolabeled contaminant present in the *flaA* probe that was observed even in the absence of MogR.

Structure of MogR Bound to its DNA Target Site

To further understand how MogR recognizes its target sequences, we crystallized the MogR truncation protein comprising residues 1–162 bound to a 15 bp DNA recognition sequence from the *flaA* promoter region. The $MogR_{1-162}$ construct was used because it was the minimal domain analyzed sufficient for DNA binding (Figures 1A and 2). The MogR:DNA crystals diffracted to a minimum Bragg spacing of 1.8 Å, and the structure was determined using single-wavelength anomalous dispersion of a seleneomethionine-labeled sample (Table S1). The structure was refined using data extending to a minimum Bragg spacing of 1.8 Å and we obtained a R_{work} and R_{free} of 21.6% and 23.2%, respectively. In our structure, we observed no electron density for MogR region 144–162, indicating that this region is flexible.

The structure contains two $MogR_{1-162}$ molecules (domain A and B) bound to the 15 bp binding site (Figure 3). The two DNA binding domains are related by a two-fold symmetry axis through the central A:T base pair (in Figure 3A this two-fold axis is indicated by a vertical dashed line). Because of this symmetry, the detailed interactions at the protein-DNA interface are identical for the two domains. The conformation of the 15 bp duplex is regular B-DNA and is bent away from the MogR dimer toward the major groove by ~52°. In the crystal, DNA duplex ends are stacked end to end to form pseudocontinuous double helices throughout the crystal. The DNA sequence is pseudopalindromic with a center of symmetry at the central A:T base pair. All base pairs of the 15 bp binding fragment, except the central base pair (A8-T8′), are engaged in MogR binding (Figures 4 and 5). The central base pair is a spacer for the half-operators and does not appear to contribute directly to sequence specificity.

The MogR DNA binding domain contains seven α helices (Figure 3, labeled α1−7 in domain A and α1′−7′ in domain B) connected by short linkers. The first three helices form an antiparallel three-helix bundle. Helix 4 forms a small dimerization interface with a buried surface area of 157A^2 (Figure S3). Helices 5–7 form a three-helix bundle DNA binding domain that contains a helix-turn-helix (HTH) motif (α6 and α7), in which α7 is the recognition helix. Helices α5 and α6 are antiparallel and helix α7 is roughly perpendicular to the axis established by these two helices. The topology of helices 5–7 is most closely related to that seen in homeodomain proteins; a Dali (Holm and Sander, 1996) search shows highest similarity to Pax6 (Xu et al., 1999), a human homeodomain protein, with an rmsd between superimposed Cα atoms from the HTH motif and the minor groove interaction loop of 3.4 Å (Z-score of 4.3).

One of the most striking features of the structure are the minor groove contacts by residues in the C-terminal loop L3 that add further binding site specificity (Figure 4B). Each half-site is recognized in the major groove by the recognition helix α7 (Figure 4A) and in the minor groove by loop L3 from the symmetry-related molecule. Therefore, the whole 15 bp recognition element, except the central A:T base pair, forms a composite binding site for binding of a MogR dimer.

DNA Binding Specificity through Major Groove Contacts by Recognition Helix α7

α6 (residues 94–105) and α7 (residues 113–125), separated by a longer than usual seven-residue turn, form a HTH unit. Docking of α7 is stabilized by the phosphate contacts from the N-terminal portion of α6 and from the C-terminal portion of the loop L3 (Figures 3A and 5). Direct base contacts by residues in the recognition helix α7 are responsible for specific DNA sequence recognition. These contacts include (from the N to the C terminus): (1) contacts between Ser114 and the N7 of adenine 4′; (2) contacts between Gln117 and the O4 of thymine 3 and thymine 2; (3) bidentate contacts between Asn118 and N7 of adenine 5′ and N6 of adenine 4′; and (4) van der Waals contacts between Tyr121 and the methyl group of thymines 2 and 3 (Figures 4A and 5). In addition, there are a number of water-mediated contacts (Figure 5). These contacts account for conservation of the flanking A:T base pairs in MogR binding sites (5′ TTTT NNNNNAAAA 3′) and are consistent with the available biochemical data.

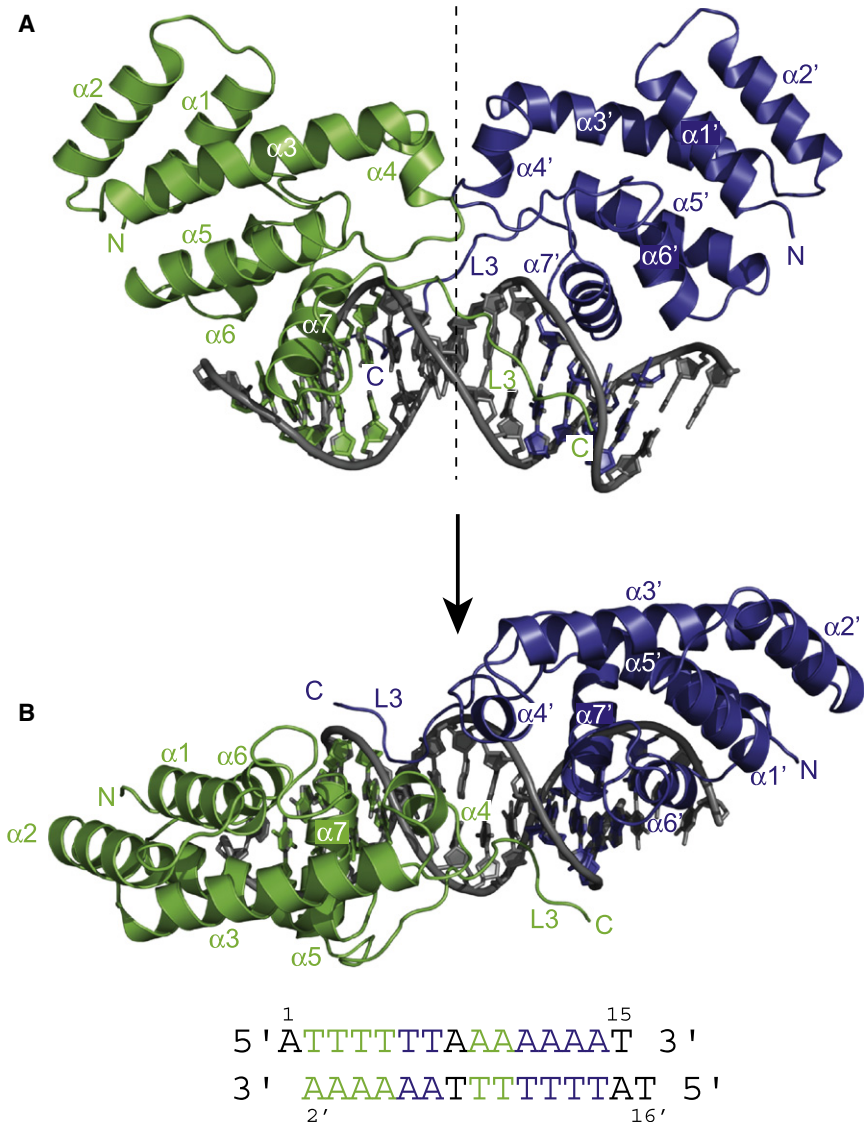

Figure 3. Structure of the MogR-DNA Complex

Domain A is shown in green and domain B in blue. Two views of the complex, related by 90°, showing the overlapping DNA contacts of both domains.

(A) Side view of the complex. Loop L3 of domain A inserts into the minor groove opposite recognition helix α7′ of domain B and vice versa. The two-fold symmetry axis is shown as a dashed line.

(B) Top view of the complex. The sequence of the cocrystallized DNA oligonucleotide is shown at the bottom of the figure and corresponds to the region −33 to −19 nucleotides from the start site of transcription of the *flaA* promoter. The core-binding site for each domain is shown with the corresponding color and the nucleotide numbering scheme is indicated. The final model contains MogR amino acids 4–143. All figures were generated with PyMol.

DNA Binding Specificity through Minor Groove Contacts by Loop L3

The extended polypeptide loop L3 (residues 127–143) lies in the minor groove and makes extensive contacts over a 5 bp region of the DNA. These contacts are on the opposite face of the DNA helix where the recognition helix α7 of the symmetry-related molecule makes major groove contacts and include: (1) van der Waals contacts between Pro138 and C2 of adenine 9 and O2 of the paired thymine 9′; (2) a hydrogen bond between the backbone amide of Gly139 and the O2 of thymine 9′; (3) a hydrogen bond between the backbone amide of Arg140 and N3 of adenine 11; (4) a hydrogen bond between the guanidinium side chain of Arg140 and the O2 of thymine 11′; and (5) a van der Waals contact between the guanidinium side chain of Arg140 to the N3 of adenine 12 (Figures 4B and 5). In addition, Arg140 has water-mediated contacts in the minor groove (Figure 5). These minor groove interactions are important for MogR specificity, as they would restrict against the presence of G:C base pairs at these positions. The exocyclic N2 amino group in the minor groove of G:C base pairs would lead to steric repulsion of the minor groove loop. Therefore we propose that the consensus MogR binding site is 5′ TTTTWWNWWA AAA 3′ (W = A or T).

DNA Conformation

The MogR recognition site contains two A-tract half-sites separated by a central TpA base pair step (residues T7A8 in the binding site). We used the programs 3DNA (Lu and Olson, 2003) and Madbend (Strahs and Schlick, 2000) to analyze the DNA helical parameters. The MogR binding site has a relatively standard B-DNA conformation with an average helical twist of 35.8° (10.0 bp/turn). The three base pairs flanking the central TpA step all have high degrees of propeller twist (>−20°) and engage in bifurcated hydrogen bonds. The minor groove is narrow at both ends of the duplex and widens toward the center with a maximum at the central TpA base pair step (Figure S4A). The central six base pairs all exhibit positive roll angles with a maximum of ~10° at the central TpA step (Figure S4B). This positive roll of consecutive base pairs results in a global bend of ~52° toward the major groove. Although the DNA duplexes stack to form a pseudocontinuous helix in the crystal, we do not believe that the overall DNA conformation is significantly influenced by packing distortions. The HTH motifs of both domains have canonical phosphate backbone contacts and the close apposition of helix α4 dictates their relative orientation. Therefore, the configuration of the two MogR monomers requires the DNA bend that we observe.

DISCUSSION

The repression of flagellar gene transcription by MogR is an important component of the regulatory network-governing

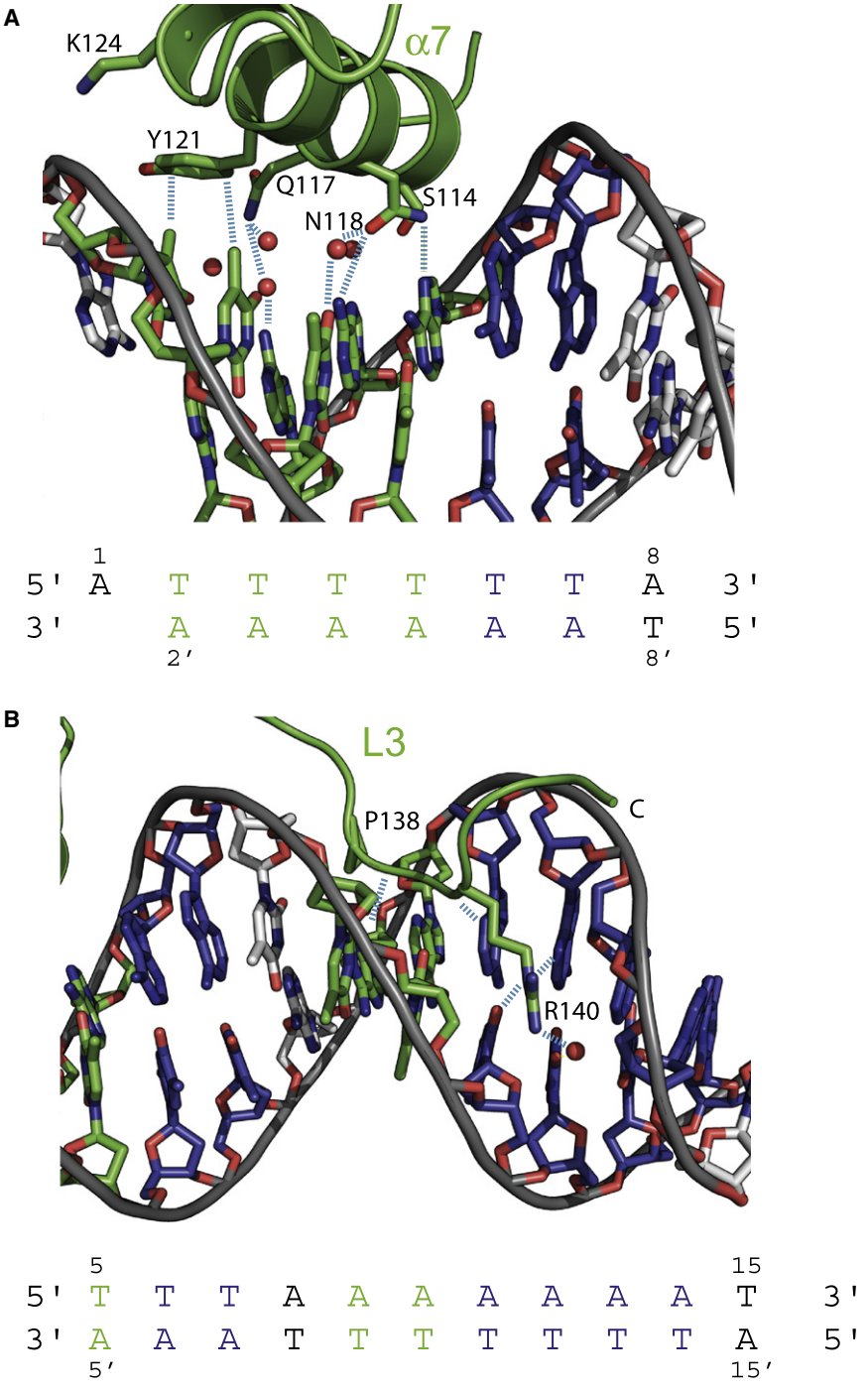

Figure 4. DNA Recognition by MogR

(A) Major groove interactions by recognition helix α7. Atoms are represented as sticks (green, carbon; red, oxygen; blue, nitrogen) Water molecules are represented as red spheres. Hydrogen bonds are shown as dashed blue lines. DNA backbone interactions are omitted for clarity. The recognition half-site is shown below with the core-binding residues highlighted in the matching color.

(B) Minor groove interactions by loop L3. Atoms are represented as sticks using the same color code as in (A).

adaptation of *L. monocytogenes* to infection of the host. Previous studies revealed that MogR downregulates flagellar gene transcription by binding to 5′ TTTTNNNNNAAAA 3′ sequences enriched within flagellar gene promoter regions (Shen and Higgins, 2006). The precise mechanism by which MogR recognizes its AT-rich binding sequences and distinguishes among the many nonspecific sites in the AT-rich genome of *L. monocytogenes* remained unknown. In this study, we defined the structural elements within MogR that permit stringent recognition of target sequences to mediate repression of flagellar gene expression. Amino acid sequence alignments indicate the presence of MogR homologs in other bacterial species (Figure 6). Residues making important DNA backbone and base-specific contacts in helix α7 and loop L3 are entirely conserved. Therefore, a similar DNA binding mechanism and mode of flagellar gene regulation through MogR-like master regulators is likely to occur in these bacterial species.

MogR Domain Organization

Our domain analysis of MogR indicated that a truncated protein containing amino acids 1–162 was competent for DNA binding, whereas a truncation containing residues 1–140 was not (Figure 2C). The MogR structure indicates that residues 140–143 are part of loop L3 involved in minor groove binding. The failure of $MogR_{1-140}$ to bind DNA is consistent with an important role of this minor groove binding loop. We crystallized the MogR truncation containing amino acids 1–162 and observed that MogR residues 144–162 are disordered, suggesting that this region is a flexible linker. Analytical ultracentrifugation experiments showed that the crystallized $MogR_{1-162}$ is monomeric in solution (data not shown).

The region comprising MogR residues 218–263 contains a putative leucine zipper region (Figure 1). The presence of this region significantly increases the apparent size of MogR by gel filtration analyses, suggesting that it is required for dimerization/oligomerization (Figure 1C). MogR mutants lacking this region still bound *flaA* promoter region DNA in vitro, but with lower apparent affinity (Figure 2 and Figure S1), again implying that this region is involved in dimerization/oligomerization. Alternatively, it is possible that full-length MogR adopts an elongated shape in the presence of the C-terminal region leading to reduced mobility in a gel filtration column. Unfortunately, the solution properties of full-length MogR have not allowed us to

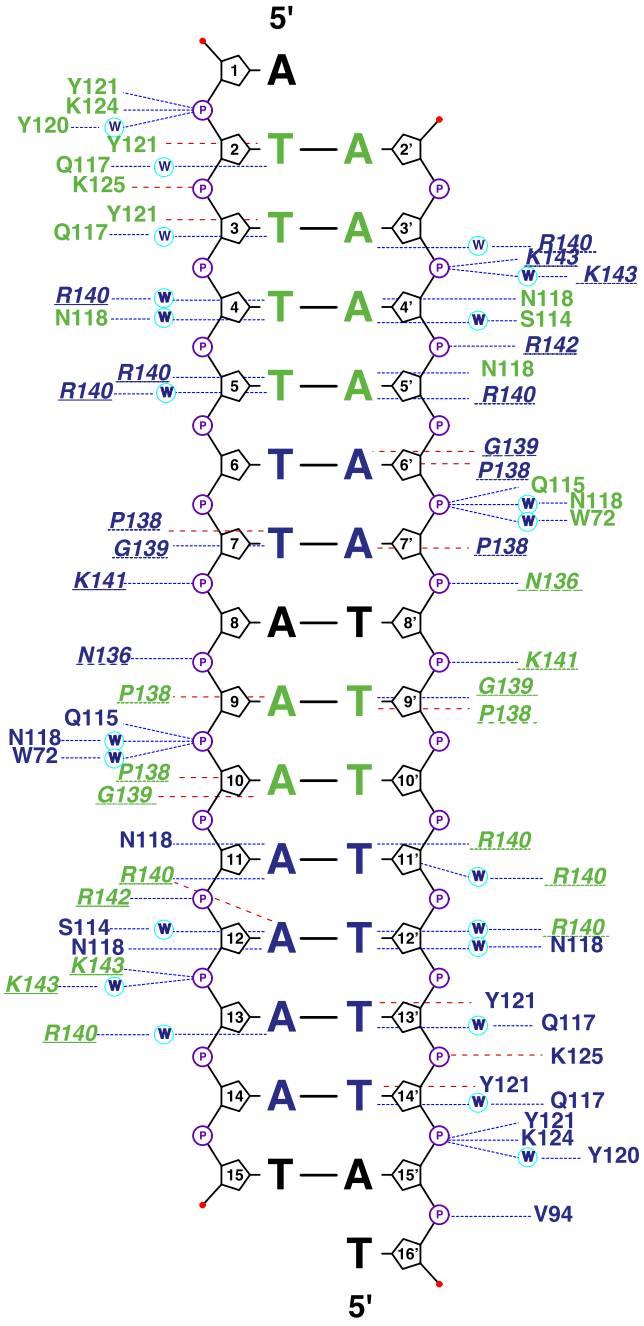

Figure 5. Schematic Diagram of Protein-DNA Contacts

Residues shown in green are from domain A and those indicated in blue are from domain B. The core-binding site for each domain is indicated in green or blue. Circles labeled P denote phosphate atoms. Circles labeled W denote water molecules. Blue dotted lines indicate hydrogen bonds and red dashed lines van der Waals contacts. Note the overlapping contacts in each half-site through minor groove interactions of loop L3 (residues 136–143, shown in italics and underlined). Figure generated using NUCPLOT (Luscombe et al., 1997).

measure its oligomerization status more rigorously. A MogR mutant lacking the leucine zipper domain ($MogR_{1-220}$) was destabilized in *L. monocytogenes* and thus failed to repress flagellar gene transcription compared to MogR proteins harboring the leucine zipper domain (Figure 1 and Figure S2). Although we were unable to rigorously assess the function of the leucine zipper domain in vivo or in vitro, we favor the model that the C-terminal region is required for MogR dimerization and higher order interaction on tandem MogR binding sites.

Basis for DNA Recognition by MogR

MogR recognizes its DNA target sequence through a HTH motif (Figure 3). As in most other HTH-containing proteins, MogR functions as a homodimer with each motif binding one half-site of the symmetry-related target DNA site. Major groove interactions of α7 of the HTH motif account for the strict conservation of the flanking T and A base pairs in the binding site 5′ TTTTNNN NNAAAA 3′ (Figure 4A). One of most striking features are the extensive minor groove interactions by loop L3 that partially overlap with the binding site of the symmetry-related molecule (Figure 4B). Minor groove contacts by an extended polypeptide chain have been seen in other complexes, such as homeodomains and interferon response factors (Kissinger et al., 1990; Panne et al., 2004; Wolberger et al., 1991; Xu et al., 1999). MogR is unusual because the major and minor groove interactions of each domain are mutually overlapping and therefore the binding site is a composite site for recognition of a MogR dimer. This "oversampling" of each half-site is unusual among HTH-containing proteins and clearly increases site specificity. One possibility is that this oversampling has evolved to distinguish among closely related sites in the AT-rich genome. The minor groove interactions also account for evolutionary conservation of the stretch of A:T base pairs in the center of MogR binding sites, as they are additional determinants of site specificity. We propose that the consensus MogR site is 5′ TTTTWW NWWAAAA 3′ (W = A or T). This sequence specificity also explains our previous observation that the central nucleotides are more permissive to mutation provided that the substitutions are conservative (A→T or T→A) (Shen and Higgins, 2006). The refinement of MogR's recognition specificity reduces the number of possible MogR binding sites in the *L. monocytogenes* genome from 534 to 92 (within 300 bp upstream of gene start codons; 0 mismatch permitted) and from 4884 to 1404 (1 mismatch permitted).

An emerging theme in protein-DNA recognition is that minor groove interactions can be important to discriminate among closely related binding sites. In the homeodomain protein Scr, a loop inserts into the minor groove only in the functionally relevant target site (Joshi et al., 2007). Specificity is obtained by recognizing a sequence-dependent DNA structure rather than by direct DNA sequence readout. By analogy to Scr, we suggest that (1) the major groove interactions target MogR to AT-rich binding sites, (2) the additional minor groove interactions are important to discriminate among closely related target sites, and (3) the intrinsic DNA structure of the target site is an important parameter in site recognition. Further experiments are required to analyze MogR binding site requirements in more detail.

Higher Order Assembly of MogR Dimers on Promoter Region DNA

Seventy-two genes contain at least one of the 92 consensus 5′ T TTTWWNWWAAAA 3′ MogR binding sites within their promoter

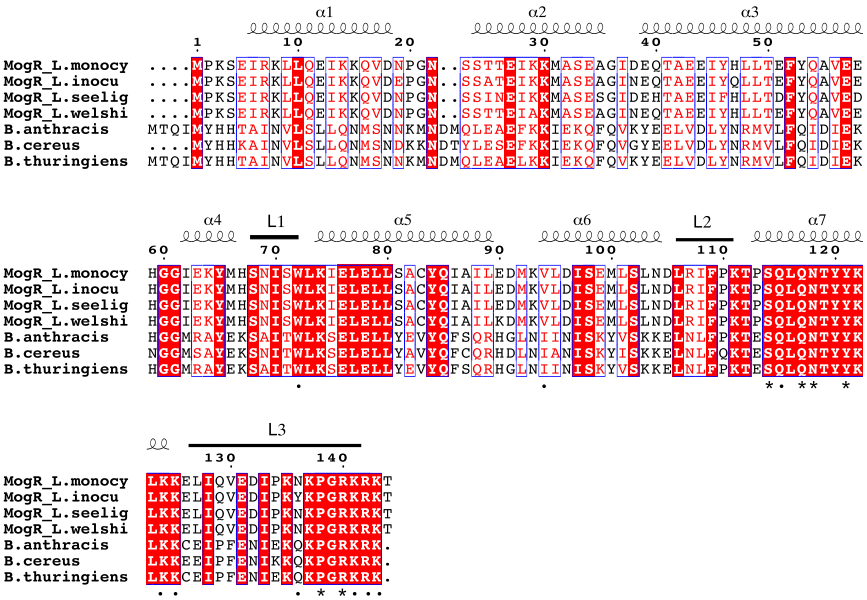

Figure 6. Multiple Sequence Alignment of MogR from *Listeria monocytogenes* (MogR_L.monocy) with homologs from *Listeria innocua* (MogR_L.inocu), *Listeria seeligeri* (MogR_L.seelig), *Listeria welshimeri* (MogR_L.welshi), *Bacillus anthracis* (hypothetical protein BAS_1573), *Bacillus cereus* (hypothetical protein BCE_1370), and *Bacillus thuringiensis* (B.thuringiens; hypothetical protein BT_1545)

Identical residues are shown with a red background and similar residues with red letters. Secondary structure elements of MogR are shown above the amino acid sequences. Residues making DNA backbone contacts are marked with a solid dot and those making base specific DNA contacts are marked with an asterisk. The high sequence conservation among MogR homologs suggests that the MogR binding mechanism is conserved among these species. The sequence alignment was performed using ClustalW (Thompson et al., 1994) and the figure was generated using Espript (Gouet et al., 2003).

regions (Table S2) and all of the known MogR-regulated promoters contain this sequence. Strikingly, promoters of MogR-regulated flagellar genes usually contain pairs of MogR binding sites that are spaced one, two, or three helical turns apart (Table S2) (Shen and Higgins, 2006). At least two MogR binding sites are required for both high affinity binding in vitro and repressor activity in vivo, another feature of MogR binding that presumably helps prevent spurious repression in the context of the AT-rich *L. monocytogenes* genome (Gründling et al., 2004; Shen and Higgins, 2006). This arrangement suggests that two MogR dimers further interact on the same face of the DNA helix to form a higher order (tetrameric) assembly. Since a single MogR dimer bends the target site by ~52°, the cumulative bend of two such sites in phase would be at least ~104°. Within the *flaA* promoter region, two MogR binding sites flank the −35 region; therefore it is likely that the higher order assembly leads to steric occlusion of the promoter as suggested previously (Shen and Higgins, 2006). In most MogR-regulated promoters, one of the MogR binding sites corresponds to a MogR consensus sequence, whereas the others are nonconsensus binding sites. We note that graded affinity for tandem operators and cooperative binding to operator sequences are central components of the mechanistic function of well studied phage and Lac repressors (Müller-Hill, 1996; Ptashne, 1986; Stayrook et al., 2008) and thus are likely to also play a key role in MogR-mediated gene regulation.

A-Tracts in Protein-DNA Recognition

Consecutive ApA, TpT, or ApT base pair steps are known to result in a narrow minor groove due to negative propeller twisting that is stabilized by interbase pair interactions in the major groove (Crothers and Shakked, 1999). In contrast, TpA base pair steps do not form these interactions and tend to widen the minor groove (Stefl et al., 2004). Hence, A-tract sequences are defined as four or more consecutive A:T base pairs without a TpA step (Hud and Plavec, 2003). By that definition, the *flaA* promoter region binding site consists of two consecutive A-tract elements disrupted by a central TpA step, 5′ ATTTTTTAAAA AAAT 3′ (TpA underlined). Considering directionality, the two A-tracts in the site are oriented "head-to-head" with their 5′ ends in the center of the binding site (at the TpA base pair step). A-tracts in the 5′→3′ direction usually exhibit a narrow minor groove (Crothers and Shakked, 1999; Hud and Plavec, 2003). Consistent with this prediction, we observe that the minor groove is narrow at both ends of the site and symmetrically widens toward the center with a maximum at the central TpA base pair step (Figure S4A). In order to distinguish between intrinsic and MogR-induced structural features of the DNA target site, we compared our structure to the DNA duplex containing a T4A4 A-tract (5′CGTTTTAAAACG 3′) as studied by solution nuclear magnetic resonance (NMR) (Stefl et al., 2004). In accord with the predictions, the NMR structure also exhibits narrow minor grooves toward the ends of the duplex and a maximum width at the central TpA step (Figure S4A). Therefore, we assume that the minor groove geometry as seen in our MogR structure is an intrinsic structural property of the DNA sequence. In the MogR structure, Arg140 inserts at a position where the minor groove is very narrow. It is known that electrostatic focusing in such narrowed grooves generates a negative electrostatic potential, which can provide the anchoring point for the positively charged arginines (Honig and Nicholls, 1995; Joshi et al., 2007).

The central base pairs exhibit large deviations in local helical parameters, which are important for the overall conformation of the site (Figure S4B). The smooth ~52° bend in the MogR DNA is due to consecutive positive roll angles in the central six nucleotide base pair steps with a maximum at the TpA step (Figure S4B). We have examined the role of individual base pair steps in the binding site by using quantitative multiple fluorescence relative affinity binding studies (Man and Stormo, 2001). We observe that a GC base pair at central position 8 does not inhibit binding significantly. However, GC base pairs at positions 7 and 9 or at 6 and 10 significantly decrease relative binding affinity (Figure S5). Whereas the N2 amide of guanine in the minor

groove can be tolerated at central position 8, at positions 6, 7, 9, and 10, they directly interfere with binding of minor groove loop L3. We also examined the role of TpA base steps in the center of the binding site. We introduced a TpA step at positions 7 and 9. TpA steps are predicted to not interfere directly with minor groove binding but to affect minor groove geometry. We observed that TpA steps at positions 7 and 9 significantly reduced binding affinity. Our interpretation is that sequence-dependent conformation of the minor groove is an important parameter for site recognition and that MogR affinity is modulated by the intrinsic structure of the central nucleotides (positions 6–10). In other protein-DNA complexes, noncontacted bases are frequently important determinants of binding affinity (Anderson et al., 1987; Hizver et al., 2001; Otwinowski et al., 1988). One explanation is that DNA binding proteins not only read the DNA sequence but also sequence-dependent DNA structure. Such sequence-dependent conformational characteristics of DNA binding sites and their ''indirect readout'' seem to explain in part why there is no simple ''code'' in protein-DNA recognition (Harrison, 2007).

EXPERIMENTAL PROCEDURES

Details regarding the construction of bacterial strains, western blot, gel mobility shift, and DNase I footprinting analyses used in this study are described in the Supplemental Experimental Procedures.

Purification of His_6-Tagged MogR Proteins

Sixteen to twenty hour cultures of pET29b-$MogR_x$ variants in *E. coli* BL21(DE3) (Novagen) were diluted 1:1000 into 250 ml of 2YT media and grown at 37°C shaking until an OD of ~0.6 was reached (approximately 3.5 hr). IPTG was added at 350 μM and cultures were induced for another 4 hr at 25°C. Cultures were pelleted and resuspended in 10 ml of LIB buffer (100 mM Tris [pH 7.5], 15 mM imidazole, 500 mM NaCl, and 10% glycerol) supplemented with 2 mM β-mercaptoethanol, 1 mg/ml lysozyme, and Complete EDTA protease inhibitor mixture (Roche) and lysed by sonication. The lysate was cleared by pelleting at 15,500 × g for a minimum of 30 min. Two-hundred and fifty microliters of Ni-NTA agarose beads (QIAGEN) was added to the cleared lysate and incubated with rotation for a minimum of 2 hr at 4°C. Samples were pelleted at 1800 × g for 1 min and washed three times in 4 ml of LIB buffer. His_6-tagged MogR variants were eluted by adding 600 μl of HIB buffer (100 mM Tris [pH 7.5], 350 mM imidazole, 500 mM NaCl, and 10% glycerol) and incubating with rotation for 15 min at 4°C.

For purification of His_6-tagged $MogR_{1-162}$ used in crystallization studies, 16–20 hr cultures of the appropriate *E. coli* strain were diluted 1:1000 in 5 L 2YT media and grown shaking at 37°C. When an OD_{600} of 0.6 was reached, IPTG was added at 350 μM and cultures were grown for 4 hr at a reduced temperature of 30°C. Cultures were pelleted, resuspended in 25 ml lysis buffer (500 mM NaCl, 100 mM Tris [pH 7.5], 15 mM imidazole, and 10% glycerol) and flash frozen in liquid nitrogen. Lysates were thawed and then lysed by sonication and cleared by pelleting at 15,000 × g for 30 min. His_6-tagged $MogR_{1-162}$ was batch affinity purified by incubating the lysates with 1.5 ml Ni-NTA agarose beads (QIAGEN) shaking for 1 hr at 4°C. Selenomethionine-substituted $MogR_{1-162}$ was produced by metabolic labeling in *E. coli* BL21(DE3) grown in minimal M9 medium supplemented with essential vitamins, nucleotide bases, and amino acids, with Met replaced by SeMet (procedure modified from that of Van Duyne et al. [1993]). Selenomethionine incorporation was verified by MALDI-TOF/TOF mass spectroscopic analysis after trypsin digestion.

Gel Filtration Analyses

Analytical gel filtration was carried out in an AKTA FPLC unit using a HiPrep 16/60 Sephacryl S-300 HR (GE Healthcare). Five-hundred microliters of affinity-purified protein was injected into a column equilibrated with 500 mM NaCl, 100 mM Tris (pH 7.5), and 10% glycerol with 0.5 mM TCEP. Elution was tracked by absorbance at 280 nm at a flow rate of 0.5 ml/min. The column was calibrated using the following standards (Bio-Rad): thyroglobulin (670 kDa), γ-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa), and vitamin B12 (1.35 kDa). The calibration curve was log M_r = 7.46 − 0.04V_e (correlation coefficient >0.99). The elution volume (V_e) was assigned using AKTA PrimeViewer software (GE Healthcare).

Preparation of DNA Used in Crystallization

Commercially synthesized DNA oligonucleotides were resuspended in 10 mM NH_4CO_3 (pH 8.0) at a final concentration of 50 μM and annealed. A series of six different duplexes ranging in size from 15 to 25 bp, all containing a single A/T overhang, was tested in crystallization trials with the 15-mer yielding well diffracting crystals.

Crystallization and Data Collection

The complex of MogR and DNA was prepared by mixing $MogR_{1-162}$ with DNA at a final concentration of 190 μM:200 μM in water. Crystals suitable for X-ray structural analysis were obtained with the hanging drop vapor diffusion method. A volume of 1 μl of the MogR:DNA complex solution was mixed with 1 μl precipitant solution containing 100 mM 2-(N-morpholino)ethanesulfonic acid (MES) (pH 6.0), 10%–15% (w/v) PEG 4000, and equilibrated against 1 ml well solution. The first crystals appeared ~20 min after setup of the drops. The largest crystals grew as rods to a maximum size of 0.2 × 0.05 × 0.05 mm^3 at 21°C. Crystals were stable in a cryoprotectant buffer containing 50% (v/v) well solution and 50% (v/v) 100 mM MES (pH 6.0), 11% (w/v) PEG 4000, and 50% glycerol. The native and derivative data sets were obtained by flash freezing the crystals in liquid nitrogen, and data were collected under cryogenic conditions (100 K). We collected native and derivative diffraction data at the Argonne National Laboratory Advanced Photon Source, beamline ID24, under a nitrogen gas stream at 100 K, using a wavelength of 0.979 Å. We processed the data with HKL2000 (Otwinowski and Minor, 1997) (Table S1).

Structure Determination and Refinement

The crystals belong to the space group $P2_1$ with the unit cell dimensions a = 48.34 Å, b = 93.59 Å, c = 60.76 Å, and d = 113.40 Å. Initial phases were calculated using the SHELX suite as implemented in HKL2MAP (Pape and Schneider, 2004). Eight high-occupancy sites were found resulting in an initial correlation coefficient of 0.41 for the best solution. After density modification, the map was of sufficient quality for automated model building by ARP/wARP (Perrakis et al., 1999), as implemented in the CCP4 package (CCP4, 1994). The initial model contained 98 residues of MogR (68% of the final model). DNA was built in O (Jones et al., 1991). Structure refinement involved the rebuilding of several loops and extension of both polypeptides toward the N and C termini and was performed with Coot (Emsley and Cowtan, 2004). The structure was refined with multiple rounds of CNS 1.1 (Brunger et al., 1998) and model building in Coot. We used noncrystallographic symmetry restraints between domains A and B of MogR that were released in the final cycle of refinement. The Ramachandran plot for each domain indicated 93% of residues in the most favored region (119/8/1/0 [favored/allowed/generously allowed/disallowed]). The local DNA helical parameters were calculated using 3DNA (Lu and Olson, 2003) and the global bend angles were obtained using Madbend (Strahs and Schlick, 2000).

ACCESSION NUMBERS

Coordinates and structure factors have been deposited in the RSCB Protein Data Bank with accession code 3FDQ.

SUPPLEMENTAL DATA

Supplemental Data include Supplemental Experimental Procedures, seven tables, and five figures and can be found with this article online at http://www.cell.com/structure/supplemental/S0969-2126(09)00154-3.

ACKNOWLEDGMENTS

We would like to thank Stephen C. Harrison for support and comments on the manuscript; Ann Hochschild for comments on the manuscript, and E. Settembre and A. Schmidt for help with data collection. Data was collected at beamline ID-24 at the Advanced Photon Source of Argonne National Laboratory. This work was supported by U.S. Public Health Service grant AI53669 from the National Institutes of Health and grant MCB-0718559 from the National Science Foundation (D.E.H.). A.S. was a recipient of a Howard Hughes Medical Institute predoctoral fellowship award.

Received: October 23, 2008
Revised: February 19, 2009
Accepted: February 19, 2009
Published: May 12, 2009

REFERENCES

Andersen-Nissen, E., Smith, K.D., Strobe, K.L., Barrett, S.L., Cookson, B.T., Logan, S.M., and Aderem, A. (2005). Evasion of Toll-like receptor 5 by flagellated bacteria. Proc. Natl. Acad. Sci. USA *102*, 9247–9252.

Anderson, J.E., Ptashne, M., and Harrison, S.C. (1987). Structure of the repressor-operator complex of bacteriophage 434. Nature *326*, 846–852.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

CCP4. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Crothers, D.M., and Shakked, Z. (1999). DNA bending by adenine-thymine tracts (London: Oxford University Press).

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Feuillet, V., Medjane, S., Mondor, I., Demaria, O., Pagni, P.P., Galan, J.E., Flavell, R.A., and Alexopoulou, L. (2006). Involvement of Toll-like receptor 5 in the recognition of flagellated bacteria. Proc. Natl. Acad. Sci. USA *103*, 12487–12492.

Gouet, P., Robert, X., and Courcelle, E. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res. *31*, 3320–3323.

Gründling, A., Burrack, L.S., Bouwer, H.G., and Higgins, D.E. (2004). *Listeria monocytogenes* regulates flagellar motility gene expression through MogR, a transcriptional repressor required for virulence. Proc. Natl. Acad. Sci. USA *101*, 12318–12323.

Harrison, S.C. (2007). Three-dimensional intricacies in protein-DNA recognition and transcriptional control. Nat. Struct. Mol. Biol. *14*, 1118–1119.

Hayashi, F., Smith, K.D., Ozinsky, A., Hawn, T.R., Yi, E.C., Goodlett, D.R., Eng, J.K., Akira, S., Underhill, D.M., and Aderem, A. (2001). The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. Nature *410*, 1099–1103.

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. USA *98*, 8490–8495.

Holm, L., and Sander, C. (1996). Mapping the protein universe. Science *273*, 595–603.

Honig, B., and Nicholls, A. (1995). Classical electrostatics in biology and chemistry. Science *268*, 1144–1149.

Hud, N.V., and Plavec, J. (2003). A unified model for the origin of DNA sequence-directed curvature. Biopolymers *69*, 144–158.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Joshi, R., Passner, J.M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M.A., Jacob, V., Aggarwal, A.K., Honig, B., and Mann, R.S. (2007). Functional specificity of a Hox protein mediated by the recognition of minor groove structure. Cell *131*, 530–543.

Kissinger, C.R., Liu, B.S., Martin-Blanco, E., Kornberg, T.B., and Pabo, C.O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions. Cell *63*, 579–590.

Lu, X.J., and Olson, W.K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res. *31*, 5108–5121.

Luscombe, N.M., Laskowski, R.A., and Thornton, J.M. (1997). NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. Nucleic Acids Res. *25*, 4940–4945.

Macnab, R.M. (2004). Type III flagellar protein export and flagellar assembly. Biochim. Biophys. Acta *1694*, 207–217.

Man, T.K., and Stormo, G.D. (2001). Non-independence of Mnt repressor-operator interaction determined by a new quantitative multiple fluorescence relative affinity (QuMFRA) assay. Nucleic Acids Res. *29*, 2471–2478.

Müller-Hill, B. (1996). The lac Operon (Berlin: DeGruyter).

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

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.

Panne, D., Maniatis, T., and Harrison, S.C. (2004). Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-beta enhancer. EMBO J. *23*, 4384–4393.

Pape, T., and Schneider, T.R. (2004). HKL2MAP: a graphical user interface for phasing with SHELX programs. J. Appl. Cryst. *37*, 843–844.

Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. *6*, 458–463.

Ptashne, M. (1986). A Genetic Switch (Cambridge, MA: Cell Press).

Ramos, H.C., Rumbo, M., and Sirard, J.C. (2004). Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. Trends Microbiol. *12*, 509–517.

Shen, A., and Higgins, D.E. (2006). The MogR transcriptional repressor regulates nonhierarchal expression of flagellar motility genes and virulence in *Listeria* monocytogenes. PLoS Pathog. *2*, e30.

Stayrook, S., Jaru-Ampornpan, P., Ni, J., Hochschild, A., and Lewis, M. (2008). Crystal structure of the λ repressor and a model for pairwise cooperative operator binding. Nature *452*, 1022–1026.

Stefl, R., Wu, H., Ravindranathan, S., Sklenar, V., and Feigon, J. (2004). DNA A-tract bending in three dimensions: solving the dA4T4 vs. dT4A4 conundrum. Proc. Natl. Acad. Sci. USA *101*, 1177–1182.

Strahs, D., and Schlick, T. (2000). A-Tract bending: insights into experimental structures by computational models. J. Mol. Biol. *301*, 643–663.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.

Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., and Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J. Mol. Biol. *229*, 105–124.

Van Houdt, R., and Michiels, C.W. (2005). Role of bacterial cell surface structures in *Escherichia coli* biofilm formation. Res. Microbiol. *156*, 626–633.

Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D., and Pabo, C.O. (1991). Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell *67*, 517–528.

Xu, H.E., Rould, M.A., Xu, W., Epstein, J.A., Maas, R.L., and Pabo, C.O. (1999). Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. Genes Dev. *13*, 1263–1275.