Plasticity in Repressor-DNA Interactions Neutralizes Loss of Symmetry in Bipartite Operators*

Deepti Jain1‡‡, Naveen Narayanan§¶, and Deepak T. Nair††

From the ‡Transcription Regulation Lab and the §Genomic Integrity and Plasticity Lab, Regional Centre for Biotechnology, NCR Biotech Science Cluster, 3rd Milestone, Faridabad-Gurgaon Expressway, Bhankri Village, Faridabad 121001, the †National Centre for Biological Sciences (NCBS-TIFR), UAS-GVKV Campus, Bellary Road, Bangalore 560065, and 1Manipal University, Manipal, 576104 Karnataka, India

Transcription factor-DNA interactions are central to gene regulation. Many transcription factors regulate multiple target genes and can bind sequences that do not conform strictly to the consensus. To understand the structural mechanism utilized by the transcription regulators to bind diverse target sequences, we have employed the repressor AraR from Bacillus subtilis as a model system. AraR is known to bind to eight different operator sites in the bacterial genome. Although there are differences in the sequences of four of these operators, ORE1, ORX1, ORA1, and ORR3, the AraR-DNA binding domain (AraR-DBD) as well as full-length AraR unexpectedly binds to each of these sequences with similar affinities as measured by fluorescence anisotropy experiments. We have determined crystal structures of AraR-DBD in complex with two different natural operators ORE1 and ORX1 up to 2.07 and 1.97 Å resolution, respectively. These structures were compared with the previously reported structures of AraR-DBD bound to two other natural operators (ORA1 and ORR3). Interactions of two molecules of AraR-DBD with the symmetric operator, ORE1, are identical, but their interaction with the non-symmetric operator ORX1 results in breakdown of the symmetry in protein-DNA interactions. The novel interactions observed are accompanied by local conformational change in the DNA. ChIP-sequencing (ChIP-Seq) data on other transcription factors have shown that they can bind to diverse targets, and hence the plasticity exhibited by AraR may be a general phenomenon. The ability of transcription factors to form alternate interactions may be important for employment in new functions and evolution of novel regulatory circuits.

The binding of proteins to cognate DNA sequences with high specificity is critical for a number of cellular processes. Among such proteins, the ability to detect and bind target sequences present at different locations in the whole genome is especially crucial for the function of transcription factors. This ability enables the transcriptional regulatory pathways to function with remarkable precision and sensitivity. The specificity of transcription factor for its cognate DNA sequence is largely achieved due to non-covalent interactions between the amino acid side chains and the functional groups present on the DNA bases in the major and minor groove (1). In addition, the shape of the DNA also plays a significant role in specific protein-DNA recognition (2–5). The availability of transcription factor binding data spanning entire genomes has led to identification of new targets for transcription factors. It has been observed that a single transcription factor can bind to several different sites throughout the genome. Sequence deviations have been seen in target sites that are recognized by the same transcription factor. Examples include glp repressors of Escherichia coli and NdgR from Streptomyces coelicolor, which recognize 13 and 19 different operator sequences, respectively (6, 7). The sequence differences suggest the possibility of variations in the mode of recognition by a single transcription factor when recognizing target sequences. A systematic structural study of the binding of a transcription modulator to naturally occurring target sequences can facilitate the understanding of the diverse structural strategies employed by the regulator for precise recognition.

AraR is the key regulatory protein of the 1-arabinose metabolism in Bacillus subtilis. AraR binds to eight different operator sequences (8–10) governing five different promoters. AraR is composed of two independent domains exhibiting different functions and belonging to different families of proteins (11, 12). The smaller N terminus domain can bind DNA (AraR-DBD) even in the absence of a C terminus domain. AraR-DBD is composed of a winged helix-turn-helix motif and belongs to the GntR family of regulators (13–15). The larger C terminus domain binds 1-arabinose and belongs to the LacI/GalR family (16, 17). AraR operators are palindromic DNA sequences, and the consensus operator is 16 bp in length (9). To understand how AraR recognizes diverse operators, we selected the ORA1, ORR3, ORE1, and ORX1 operators for structural and biochemical analysis.

DNA footprinting experiments have shown that AraR may bind these different operators with varying affinities (9). However, fluorescence anisotropy experiments unexpectedly showed that the binding affinities of full-length AraR as well as AraR-DBD to various operators are similar. We have

* This work was supported by an Innovative Young Biotechnologist Award (IYBA) grant from the Department of Biotechnology (to D. J.) and intramural funds available from the Regional Centre for Biotechnology (to D. J.) and the National Centre for Biological Sciences-Tata Institute of Fundamental Research (to D. T. N.). The authors declare that they have no conflicts of interest with the contents of this article.

The atomic coordinates and structure factors (codes 5D4R and 5D4S) have been deposited in the Protein Data Bank (http://wwpdb.org/).

‡ To whom correspondence should be addressed: Tel.: 91-0129-2848839; Fax: 91-0129-2848803; E-mail: deepti@rcb.res.in.

The abbreviations used are: DBD, DNA binding domain; r.m.s.d., root mean square deviation.

1235
Plasticity in Operator-Repressor Interactions

Previously determined the x-ray crystal structure of the N terminus domain of AraR bound to two natural operators, ORA1 and ORR3 (13). These structures reveal that the two AraR-DBD monomers are bound to each bipartite operator and that recognition of each monomer by DNA is through the core recognition motif TXG present in one half-site of the operator (where X is any base). In the present study, we report the crystal structures of AraR-DBD in complex with two other natural operators, ORE1 and ORX1. Although the overall position of the protein on the DNA is similar in the two cases, there are significant differences in the hydrogen-bonding scheme in the two complexes. A detailed analysis of the AraR-DNA interactions shows that any breakdown of symmetry in the operator sequence is compensated by novel interactions between repressor and DNA. The structural analysis reveals that sequence differences lead to distinct local conformations in the DNA that can be exploited by the transcription factor to achieve novel stabilizing interactions. Overall, the observed plasticity in interactions may be a general phenomenon that allows transcription factors to bind disparate DNA sequences without compromising the specificity.

Materials and Methods

Fluorescence Anisotropy—Fluorescence anisotropy was measured for both full-length AraR and DBD using four different operators, namely ORA1, ORE1, ORX1, and ORR3 (see Fig. 1). The 3’ end of one of the oligonucleotides in the duplex was labeled with 6-fluorescein amidite, procured from the Keck Biotechnology Resource Laboratory (Yale University). The protein in varying concentrations (0–500 nM for AraR-DBD and 0–100 nM for full length) was incubated with 1 nM of respective DNA in a buffer containing 25 mM Tris, pH 8.0, 5% glycerol, 1 mM DTT, 150 mM NaCl for 45 min at room temperature. All the reactions were carried out in triplicates, on 96-well black bottom Costar® plates. The fluorescence anisotropy was measured using a SpectraMax M5 micro plate reader (Molecular Devices) with an excitation wavelength of 492 nm and an emission wavelength of 517 nm. SpectraMax software calculates anisotropy values using I parallel (||) and I perpendicular values (⊥⊥).

Three independent measurements were recorded for each protein concentration, and average anisotropy values were calculated. Reduced anisotropy values were calculated by subtracting average anisotropy values for a particular protein concentration, and average anisotropy values were calculated without protein. The reduced anisotropy values were then divided with the maximum value of reduced anisotropy to obtain the fraction bound. This fraction bound was plotted (on the y axis) against protein concentration (on the x axis). The fraction bound data were fitted using the Hill equation to calculate the $K_d$ values. This experiment was repeated with all four operators.

AraR-DBD Cloning, Expression, and Purification—The gene segment corresponding to AraR-DBD (residues 1–68) was cloned, expressed, and purified as described earlier (13).

Nucleic Acid Preparation—Four oligonucleotides corresponding to complementary sequences for the ORE1 and ORX1 (with a T/A overhang) were purchased from Sigma and dissolved in appropriate volumes of autoclaved and filtered (0.22-μm filter) water to achieve a final concentration of 4 mM. Equimolar amounts of complementary oligonucleotides were annealed by heating to 90°C for 5 min followed by cooling to 25°C, and the final concentration of the duplex DNA was 2 mM.

Crystallization—Crystallization conditions and the cryoprotection strategy for the AraR-DBD$\text{OR}_{\text{RE1}}$ complex were identical to that for AraR-DBD$\text{OR}_{\text{RX1}}$. Complex co-crystals were obtained using vapor diffusion by mixing the duplex DNA and AraR-DBD-(1–68) in the ratio 1.2:1, with the final concentration of DNA at 0.67 mM. The mixture was incubated on ice for 20 min. The complex crystallized in solutions of PEG 8000 (20%) buffered with 0.1 mM sodium acetate (pH 4.5) containing 200 mM KCl. Crystals were cryoprotected by serial transfers using increasing concentrations of glycerol from 5 to 25% in steps of 5%. The crystals were flash-frozen in liquid nitrogen.

Structure Determination—The native data sets were collected for AraR-DBD$\text{OR}_{\text{RE1}}$ and AraR-DBD$\text{OR}_{\text{RX1}}$ at the BM14 beamline of the European Synchrotron Radiation Facility (see Table 1). The data were processed using HKL2000 followed by scaling and merging in SCALEPACK (18). The AraR-DBD$\text{OR}_{\text{RE1}}$ and AraR-DBD$\text{OR}_{\text{RX1}}$ complex structures were solved by the molecular replacement method in PHASER (19) using one molecule of AraR-DBD and double-stranded DNA as the model from the previously solved AraR-DBD$\text{OR}_{\text{RX1}}$ structure (4H0E) as the model. The second molecule of AraR-DBD was docked into the electron density, and the map was improved through iterative cycles of rigid body refinement in CNS. The DNA sequence of ORR3 was modified to match that of ORE1 and ORX1 sequences in Coot (20). This was followed by cycles of model building (Coot) refinement (CNS and PHENIX) (21, 22) and water picking (PHENIX) with constant monitoring of geometrical parameters. Final rounds of refinement were performed in PHENIX incorporating TLS restraints (23). The $R_{\text{free}}$ and $R_{\text{work}}$ converged to final values of 24.7%/19.7% and 24.1%/19.5% for the AraR-DBD$\text{OR}_{\text{RE1}}$ and AraR-DBD$\text{OR}_{\text{RX1}}$, respectively (see Table 1). For the final refined models of the two complexes, MOLPROBITY (24) revealed that 98% of the residues of the DBD are in the allowed regions of the Ramachandran plot. The structure was analyzed using CONTACT (CCP4 suite (25)). Figures were generated using PyMOL (Schrodinger Inc.).

Results

AraR Binds Diverse Sequence Operators with Similar Affinities—The sequences of the operators used in the study are shown in Fig. 1A. The oligonucleotides are numbered on the basis of the 16-bp consensus sequence (13). There is a dyad symmetry in the operator sequences such that each half-site of the oligonucleotide has ATTTGTAC as the consensus sequence and contains one recognition motif: TXG. The operator ORX1 is asymmetric because it lacks this motif in the first half-site. The spacing between the two recognition motifs is 6 bp for ORA1, ORE1, and ORX1 and is 8 bp in case of ORR3. These operators are conserved at positions 1, 6, 8–11, 15, and 16. (Fig. 1A).

We have analyzed the increase in fluorescence anisotropy on titrating the labeled oligonucleotides with either the full-length AraR or the DBD (26). The $K_a$ for the full-length AraR to vari-
ous operators is in the range 69–80 nM (Fig. 1B), and the affinity of DBD-AraR to operators is lower ($K_d$/$H_{11005}$ 247–277 nM). Based on the composition of these sequences and available data in the literature, it was expected that the affinity for these sequences will be different (9). Previously, the affinities of AraR to the various operators have been measured using DNA footprinting experiments and have shown apparent affinity constant ($K_{app}$) ranges between 40 nM and $H_{11022}$ 250 nM (9). The full-length AraR binds DNA with higher affinity as compared with DBD. Given that the sequence of ORX1 differs from the other operators in one half-site, it was surprising that the DBD bound to it with similar affinity as the other operators. This observation suggests that there may exist a structural strategy to neutralize the effect of sequence divergence on the thermodynamics of AraR-DBD association.

Overall Structure of AraR-DBDORE1 and AraR-DBDORX1—AraR-DBD-(1–68) was co-crystallized in complex with a 21-bp oligonucleotide containing the natural operators ORE1 (2.07 Å) and ORX1 (1.97 Å), and the structures have been determined by molecular replacement (Table 1). The DNA duplexes used for crystallization have a T/A overhang, and they pack head-to-tail in the crystal. There are two molecules of AraR-DBD (monomers A and B) bound to one molecule of double-stranded DNA in the asymmetric unit. Each monomer of AraR-DBD recognizes one half-site of the operator without forming a dimer (Fig. 2A and B). The DNA binding domain of AraR adopts winged helix-turn-helix architecture. The AraR-DBD has two distinct DNA binding elements. One is the helix-turn-helix motif formed by helix 2 and helix 3 of AraR-DBD that interacts with DNA bases in the major groove through the residues present at the tip of the recognition helix (helix 3). The second is the wing motif formed by the residues present in the loop connecting the two $\beta$-strands, which form base-specific interactions in the adjacent minor groove (Fig. 2A). In addition, there are several sequence-independent contacts between protein backbone atoms and/or side chains and the sugar-phosphate backbone of DNA (Fig. 3).

AraR-DBDORE1 and AraR-DBDORX1 structures can be superimposed on the AraR-DBDORA1 complex (13) with a root mean square deviation (r.m.s.d.) of 0.23 Å for 136 C$\alpha$ pairs. The r.m.s.d. values for superimposition of one monomer onto another in the ORE1 and ORX1 complexes (68 C$\alpha$ pairs) are 0.21 and 0.18 Å, respectively.

AraR-DBDORE1 Complex—ORE1 is a symmetric operator, and the recognition motif (T$X$G) is present in both half-sites (Fig. 1A). Any deviation from the T$X$G is highlighted in red. The center of symmetry of the operators is denoted by the asterisk. $B$, the binding affinities of AraR-DBD and full-length AraR for the four operators.

![Figure 1](image1.png)

**A**

| Oligonucleotide | DNA binding domain | Full length AraR |
|-----------------|--------------------|------------------|
| $K_d$ (nM)      | $R^2$              | $K_d$ (nM)       | $R^2$ |
| ORA1            | 263.6±6.2.6        | 0.98             | 73±0.8 | 0.96 |
| ORX1            | 277.0±3.09         | 0.96             | 69±0.39| 0.98 |
| ORR3            | 277.3±3.08         | 0.99             | 70.8±0.4| 0.98 |
| ORE1            | 247.6±2.2          | 0.99             | 80.2±0.28| 0.98 |

**B**

**TABLE 1**

| Data collection and refinement statistics | AraRNTDORE1 (Native)* | AraRNTDORX1 (Native)* |
|------------------------------------------|-----------------------|-----------------------|
| Wavelength (Å)                           | 1                     | 1                     |
| Space group                              | C2                    | C2                    |
| Cell (Å)                                 | 134.1, 42.4, 67.2     | 138.2, 42.5, 67.2     |
| °                                        | 90.0, 114.8, 90.0     | 90.0, 115.0           |
| Resolution (Å)                           | 2.07 (2.11-2.07)      | 1.97 (2.0-1.97)       |
| $R_{merge}$                              | 6.2 (32.7)            | 4.9 (44.8)            |
| $I$/$I_{merge}$                          | 10.0 (3.24)           | 10.6 (2.25)           |
| Completeness (%)                         | 99.2 (93.1)           | 99.8 (97.8)           |

| Refinement                                | 36-2.07               | 36-1.97               |
| Resolution (Å)                            | 21,116                | 25,149                |
| No. of Reflections                        | 19.7/24.7             | 19.5/24.1             |
| Average B-factors (Å²)                    | 32.9                  | 41.3                  |
| Protein: A                                | 26.82                 | 36.2                  |
| DNA                                       | 31.04                 | 37.3                  |
| Water                                     | 35                    | 41.2                  |
| r.m.s.d.                                  | 0.01                  | 0.02                  |
| Bond angles (°)                           | 1.18                  | 2.03                  |

$*$ Highest resolution shell is shown in parentheses.

$^b$ $R_{free}$ was calculated using 5% of data excluded from the refinement statistics.

![Figure 2](image2.png)

**FIGURE 2. Structure of AraR-DBD-operator DNA complex.** A, the structure of AraR-DBDORE1 complex. Protein is shown in ribbon representation with DNA binding elements labeled. Monomer A and B are colored yellow and green, respectively. DNA is shown in stick representation, and the 5’ and 3’ ends are labeled. B, the structure of AraR-DBDORX1. The figure is color-coded as in A.
As a result, AraR-DBD monomers A and B form similar interactions with ORE1. The complete intermolecular contacts of AraR-DBD$_{ORE1}$ complex are shown in Fig. 3A. Only the interactions of monomer A with DNA are described below.

The side chains of monomer A, Arg$^{41}$:A and Gln$^{61}$:A, form base-specific interactions with Gua$^5$ (the guanidinium group forms hydrogen bonds with the O6 and N7 atoms of Gua$^5$) and nucleotide Ade$^3$/H11032 in the DNA major and minor groove, respectively (Figs. 3A and 4A). In addition, Arg$^{41}$:A also forms a water-mediated hydrogen bond with the base Ade$^6$. His$^{42}$:A forms a water-mediated hydrogen bond with Gua$^8$ and Ade$^{10}$. Interestingly, in the AraR-DBD$_{ORE1}$ complex, His$^{42}$:B (monomer B) is present in two different conformations. In one conformation, it comes closer to DNA and forms a base-specific interaction with Gua$^9$ and a water-mediated interaction with Thy$^{10}$ (Fig. 3A). The residue Gly$^{62}$:A forms water-mediated hydrogen bonds with bases Thy$^2$ and Thy$^1$. Similar water-mediated interactions are seen for the residues of monomer B, G62:B (Thy$^{16}$, Thy$^{15}$) and R41:B (Ade$^{11}$). In addition, there are extensive interactions of protein residues with the sugar-phosphate backbone of DNA. The residues Glu$^{30}$, Gly$^{62}$, Arg$^{45}$, Tyr$^{4}$, Lys$^{4}$, Ser$^{40}$, and Thr$^{43}$ form interactions with the DNA backbone in both the monomers (Fig. 4A).

AraR-DBD$_{ORX1}$ Complex—In the case of the natural operator ORX1, the recognition motif present in the first half-site is not TXG but ACA, and consequently, the operator sequence is asymmetric (Fig. 1). As a result, AraR-DBD monomers A and B form similar interactions with ORX1. The complete intermolecular contacts of AraR-DBD$_{ORX1}$ complex are shown in Fig. 3A. Only the interactions of monomer A with DNA are described below.

The side chains of monomer A, Arg$^{41}$:A and Gln$^{61}$:A, form base-specific interactions with Gua$^5$ (the guanidinium group forms hydrogen bonds with the O6 and N7 atoms of Gua$^5$) and nucleotide Ade$^3$/H11032 in the DNA major and minor groove, respectively (Figs. 3A and 4A). In addition, Arg$^{41}$:A also forms a water-mediated hydrogen bond with the base Ade$^6$. His$^{42}$:A forms a water-mediated hydrogen bond with Gua$^8$ and Ade$^{10}$. Interestingly, in the AraR-DBD$_{ORX1}$ complex, His$^{42}$:B (monomer B) is present in two different conformations. In one conformation, it comes closer to DNA and forms a base-specific interaction with Gua$^9$ and a water-mediated interaction with Thy$^{10}$ (Fig. 3A). The residue Gly$^{62}$:A forms water-mediated hydrogen bonds with bases Thy$^2$ and Thy$^1$. Similar water-mediated interactions are seen for the residues of monomer B, G62:B (Thy$^{16}$, Thy$^{15}$) and R41:B (Ade$^{11}$). In addition, there are extensive interactions of protein residues with the sugar-phosphate backbone of DNA. The residues Glu$^{30}$, Gly$^{62}$, Arg$^{45}$, Tyr$^{4}$, Lys$^{4}$, Ser$^{40}$, and Thr$^{43}$ form interactions with the DNA backbone in both the monomers (Fig. 4A).

Another residue, His$^{42}$:A side chain, also undergoes a significant conformational change, and it forms hydrogen bonds with the N7 and O6 atoms of Gua$^8$ (Fig. 4B). This novel interaction is possible due to the local conformational change in the DNA (Fig. 5B). This interaction does not occur in the AraR-DBD$_{ORE1}$ complex as it would result in steric clash between His$^{42}$:A and Gua$^8$ (Fig. 5A). It has been seen earlier that the mutation of Gua$^8$ to Thy results in complete loss of repression activity in...
Comparison of AraR-DBD$_{ORE1}$ and AraR-DBD$_{ORX1}$ Structures with Other AraR-DBD-Operator Complexes—Binding of the recognition helix of the helix-turn-helix motif into the major groove of DNA allows for base-specific interactions with protein side chains. AraR-DBD forms very similar interactions with the DNA in AraR-DBD$_{ORE1}$, AraR-DBD$_{ORR3}$, and AraR-DBD$_{ORA1}$ complexes, as can be seen in Fig. 4, A–D. In all three structures, the side chain of Arg$^{41}$ forms specific interactions with Gua$_5$ of the TXG motif, and Gln$^{61}$ forms specific interactions with Ade$_3^\prime$, which is present opposite to Thy in the TXG motif. Also, in all three complexes, the interactions of monomer A and B with DNA are symmetric. However, the operator, ORX1, contains the altered recognition motif (ACA) in the first half-site of the operator ORX1. As a result, in this complex, the AraR-DBD monomer A makes novel contacts with DNA (Fig. 4B) that are facilitated by local conformational change in the DNA. Consequently, the interactions of AraR-DBD become asymmetric with the two half-sites of the operator ORX1.

Analysis of DNA Structure—It has been shown that the shape of DNA contributes extensively toward protein-DNA recognition (2–5). The analysis of the DNA structure using Curves+ (29) reveals that although the ORE1 operator is bent by 10° when bound to AraR-DBD, the ORX1 operator is bent by 8.5°. The electrostatic potential surface in the two complexes is shown in Fig. 6, A and B. Although there is no significant change in the charge on the interacting surface, there is a difference in the shape of the protein where it contacts the altered recognition motif in the ORX1 complex. The structure of the DNA bound to the protein was also analyzed for all four repressor-operator complexes. The widths of the major groove, minor groove, twist, and roll parameters measured for DNA are plotted in Fig. 7, A–D. As can be seen from the plot, the minor groove is significantly narrower in the region of the two recognition motifs where it is interacting with the DNA. In AraR-DBD$_{ORR3}$ complex, the two recognition motifs are separated by 8 bp as opposed to 6 bp in the other three structures, and as a result, the constriction of the minor groove is also shifted by two bases. In addition, there is a slight kink in the DNA at the central CG base pair as reflected in the decrease in the minor groove width in the region in all but the AraR-DBD$_{ORR3}$ complex. The major groove is enlarged by about 2 Å in the region of protein binding in all but the ORX1 structures as compared with the standard B-DNA values. These local changes in DNA conformation in the ORX1 structure facilitate the altered interactions observed in this complex.

Discussion

Previous data have allowed us to define the specific recognition motif (TXG) present in the two half-sites of the operator sequences recognized by the DNA binding domain of AraR (13). In addition to these observations, we have now shown that the two molecules of AraR bound to the same operator are capable of recognizing significantly different DNA sequences in the two half-sites. This is achieved by providing a novel interface in one of the two monomers to bind to the asymmetric operator ORX1, keeping the position of the protein on DNA intact. Sequence comparison of the four operators (Fig. 1) shows that position three is occupied by either Thy or Ade. The
Plasticity in Operator-Repressor Interactions

AraR-DBD utilizes the same residue, Gln61, to recognize both these bases. Another deviation from the consensus in the operator ORX1 is the presence of Ade instead of Gua at position 5. The structures show that this change results in breakdown of the symmetry of interactions, and in such cases, the protein-DNA recognition is modulated by the local DNA structure and bases outside the recognition motif.

Transcription factors often bind DNA sequences that are altered or relaxed as compared with the consensus sequence. These sites have been shown to be present in unexpected locations in the genome (30). Thus, transcription factors are capable of binding to a larger repertoire of sequences than what is suggested by consensus motif searches. In fact, ChIP-Seq and ChIP-on-chip data have identified degenerate target sites for several transcription factors such as CtrA, LexA, and FNR, which were earlier unknown (31–33). Using modified chromatin immunoprecipitation, several intergenic sites were identified to be bound in vivo by transcription factor CtrA from Caulobacter (31). These binding sites did not correspond to the consensus sequence identified by programs such as MEME and BIOPROSPECTOR upon searching the upstream sequences of cell cycle regulons. In vitro binding of CtrA was shown by DNA footprinting to the fliX promoter, which only showed a 4-bp match to the minimum 8-bp consensus. Similarly, of the 63 target sites identified by ChIP-on-chip for FNR binding, 10 sites do not match the canonical binding sites (33). ChIP-on-chip, along with transcription profiling studies in the case of AraC, a transcription factor that regulates arabinose metabolism in E. coli and Salmonella enterica, have identified targets that have no connection with arabinose metabolism (34).

It has been suggested earlier that cooperative binding by multiple transcription factor to adjacent sites reduces the requirement for the consensus (32). However, from the current study, it is clear that in addition to cooperativity, these transcription factors are endowed with the ability to tolerate the deviations from the consensus site through energetically favorable alternate interactions. The altered specificity of the transcription factor as observed here can promote the rewiring of the transcription regulatory circuits and can facilitate the evolution of new ones.

A systematic study depicting the structural basis of DNA recognition by transcription factors is lacking in literature except in a very few cases. Structures of carbon catabolite protein A (CcpA) and mammalian transcription factor NF-κB p65 (RelA) have been determined when bound to three different operator sequences in each case (35–37). The two proteins bind their cognate operators with similar affinities but use different structural strategies for interactions. The CcpA-DNA structures show that the operators display different bend angles when bound to protein facilitated by a flexible linker between the two domains of the protein. The NF-κB p65 (RelA) structure bound to asymmetric and pseudosymmetric operator shows that the hydrogen-bonding patterns at the protein-DNA interface show significant differences (37) in the two cases. Sim-

FIGURE 5. A, superimposition of the monomer A from AraR-DBD_{ORE1} (yellow) and AraR-DBD_{ORX1} (magenta) structures showing the conformation of His^{42}. B, the local distortion in the DNA near the His^{42} binding site. The DNA is shown in graphic representation, and the protein is shown in ribbon representation with the His^{42} side chain shown in stick representation. The color coding is same as in panel A.

FIGURE 6. Stereo view of the electrostatic potential surface of the DNA binding site in the ORE1 (A) and ORX1 (B) complexes. The molecular surfaces are colored according to the electrostatic potential. Positive potential is shown in blue, and negative potential is in red. The recognition motif in the second half-site is labeled in white.
Similarly, the structures of the DNA binding domain of the repressor of the phage 434 bound to OR1, OR2, and OR3 sites show extensive structural differences between the consensus and non-consensus half-sites. In the case of OR3, rearrangement of several amino acid side chains occurs for DNA binding (38–40). In the case of the human estrogen receptor, it was shown that the recognition of the non-consensus DNA occurred through rearrangement of a lysine side chain and formation of an alternate base contact (41).

Taken together, these studies, along with the present study, suggest that there are at least three approaches by which plasticity favors the binding of a transcription factor. (i) The first scenario is one in which the binding of the protein to the non-consensus half-site happens only through backbone interactions; the protein side chains do not show any base-specific interactions to the altered half-site. (ii) The second scenario is one in which the binding to the altered site occurs through utilization of exactly the same residues that would recognize the consensus half of the operator. (iii) The third scenario is one in which the protein uses different sets of side chains to interact with the non-consensus half-site and the consensus half-site. It is evident that the appearance of newer specificities is facilitated by flexibility of the DNA backbone as well as the amino acid side chains. The ability of a transcription regulator to recognize different sequences using any one of the above three strategies allows scope for evolution of new regulatory circuits as any mutation in the protein can lead to the appearance of new specificities.

Author Contributions—D. J. cloned, expressed, and purified the protein, set up crystallization experiments, solved the structures, and refined the structures. N. N. carried out anisotropy experiments. D. T. N. collected the x-ray diffraction data. D. J. and D. T. N. analyzed the structures. D. J. wrote the manuscript.

Acknowledgments—We acknowledge the help rendered by Dr. Hassan Belrhali (European Synchrotron Radiation Facility (ESRF)) and Dr. Babu Manjashetty during data collection at the BM-14 beamline of ESRF. Data collection at the BM14 beamline of ESRF (Grenoble, France) was funded by the BM14 project, a collaboration between Department of Biotechnology (Government of India), European Molecular Biology Laboratory (EMBL), and ESRF.

References
1. Mandel-Gutfreund, Y., Schueler, O., and Margalit, H. (1995) Comprehensive analysis of hydrogen bonds in regulatory protein DNA complexes: in search of common principles. *J. Mol. Biol.* 253, 370–382
2. Slattery, M., Zhou, T., Yang, L., Dantas Machado, A. C., Gordân, R., and Rohs, R. (2014) Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* 39, 381–399
3. Zhou, T., Shen, N., Yang, L., Abe, N., Horton, J., Mann, R. S., Bussemaker, H. J., Gordân, R., Rohs, R. (2015) Quantitative modeling of transcription factor binding specificities using DNA shape. *Proc. Natl. Acad. Sci. U.S.A.* 112, 4654–4659
4. Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., and Mann, R. S. (2010) Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* 79, 233–269
5. Abe, N., Dror, I., Yang, L., Slattery, M., Zhou, T., Bussemaker, H. J., Rohs, R., and Mann, R. S. (2015) Deconvolving the recognition of DNA shape from sequence. *Cell* 161, 307–318
6. Zhao, N., Oh, W., Trybul, D., Thrasher, K. S., Kingsbury, T. J., and Larson, T. J. (1994) Characterization of interaction of the glp repressor of *Escherichia coli* K-12 with single and tandem glp operator variants. *J. Bacteriol.* 176, 2393–2397

FIGURE 7. Helical parameters of the DNA in the four operator-AraR-DBD complexes. A–D, the width of the major groove (A), minor groove (B), twist (C), and roll (D) is plotted for each base pair step as calculated from CURVES + (29). The sequence of ORA1 is displayed on the x axis for reference. The recognition motif is in bold in the two half-sites. Canonical groove widths for B-DNA are shown as thin dashed lines.
Plasticity in Operator-Repressor Interactions

7. Kim, J.-N., Jeong Y., Yoo, J.S., Roe, J.-H., Cho, B.-K., and Kim, B-G., (2015) Genome-scale analysis reveals a role for NodG in the thiol oxidative stress response in *Streptomyces coelicolor*. BMC Genomics 16, 116

8. Raposo, M. P., Inácio, J. M., Mota, L. J., and de Sá-Nogueira, I. (2004) Functional domains of the *Bacillus subtilis* transcription factor AraR and identification of amino acids important for nucleoprotein complex assembly and effector binding. *J. Bacteriol.* 188, 3024–3036

9. Chen, Y. Q., Sengchanthalangsy, L. L., Hackett, A., and Ghosh, G. (2000) NF-κB (RelA) homodimer uses distinct mechanisms to recognize *O*1 through alternative side-chain conformation. *Structure* 8, 419–428

10. Mota, L. J., Sarmento, L. M., and de Sá-Nogueira, I. (2001) Control of the arabinose regulon in *Bacillus subtilis* by AraR in vivo: crucial roles of operators, cooperativity, and DNA looping. *J. Bacteriol.* 183, 4190–4201

11. Sá-Nogueira, I., and Mota, L. J. (1997) Negative regulation of t-arabinose metabolism in *Bacillus subtilis*: characterization of the ararL (arac) gene. *J. Bacteriol.* 179, 1598–1608

12. Franco, I. S., Mota, L. J., Soares, C. M., and de Sá-Nogueira, I. (2006) The arabinose repressor AraR from *Bacillus subtilis* gene. *FEMS Microbiol. Lett.* 263, 291–295

13. Weickert, M. J., and Adhya, S. (1992) A family of bacterial regulators homologous to Gal and lac repressors. *J. Biol. Chem.* 267, 15869–15874

14. Procházková, K., Cermáková, K., Pachl, P., Sieglová, I., Fábry, M., Otwinowski, Z., Rezácˇová, P. (2012) Structure of the effector-binding domain of the arabinose repressor AraR from *Bacillus subtilis*. *Acta Crystallogr. D Biol. Crystallogr.* 68, 176–185

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

16. McCoy, A. J., Frank, W. H., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674

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

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

19. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger T. C., and Zwart P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221

20. Winn, M. D., Murshudov, G. N., and Papiz, M. Z. (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* 374, 300–321

21. Chen, V. B., Arendall, W. B., 3rd, Head, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12–21

22. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* 67, 235–242

23. Heyd, T., Ma, Y., Tang, H., and Ebright, R. H. (1996) Fluorescence anisotropy: rapid, quantitative assay for protein–DNA and protein–protein interaction. *Methods Enzymol.* 274, 492–503

24. Correia, I. L., Franco, I. S., and de Sá-Nogueira, I. (2014) Tovard amino acid-base contacts in gene regulatory proteins: AraR – a case study. *PLoS ONE* 9, e111802

25. Franco, I. S., Mota, L. J., Soares, C. M., and de Sá-Nogueira, I. (2007) Probing key DNA contacts in AraR-mediated transcriptional repression of the *Bacillus subtilis* arabinose regulon. *Nucleic Acids Res.* 35, 4755–4766

26. Lavery, R., Moakher, M., Maddocks, I. H., Petkeviciute, D., and Zakrzewska, K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* 37, 2151–2157

27. Bonocora, R. P., Fitzgerald, D. M., Stringer, A. M., and Wade, J. T. (2013) Non-canonical protein–DNA interactions identified by ChiP are not artifactual. *BMC Genomics* 14, 254

28. Lau, M. T., Chen, S. L., Shapiro, L., and McAmads, H. H. (2002) Genes directly controlled by CtrA, a master regulator of the Caulobacter cell cycle. *Proc. Natl. Acad. Sci. U.S.A.* 99, 4632–4637

29. Wade, J. T., Reppas, N. B., Church, G. M., and Struhl, K. (2005) Genomic analysis of LexA binding reveals the permissive nature of the Escherichia coli genome and identifies unconventional target sites. *Genes Dev.* 19, 2619–2630

30. Grainger, D. C., Alba, H., Hurd, D., Browning, D. F., and Busby, S. J. (2007) Transcription factor distribution in *Escherichia coli* studies with FNR protein. *Nucleic Acids Res.* 35, 269–278

31. Stringer, A. M., Currenti, S., Bonocora, R. P., Baranowski, C., Petrone, B. L., Palumbo, M. J., Reilly, A. A., Zhang, Z., Erill, L., and Wade, J. T. (2014) Genome-scale analyses of *Escherichia coli* and *Salmonella enterica* AraC reveal noncanonical targets and an expanded core regulon. *J. Bacteriol.* 196, 660–671

32. Kalodimos, C. G., Bonvin, A. M. J. J., Salinas, R. K., Wechselberger, R., Boeens, R., and Kaptein, R. (2002) Plasticity in protein–DNA recognition: lac repressor interacts with its natural operator *O*1 through alternative conformations of its DNA-binding domain. *EMBO J.* 21, 2866–2876

33. Schumacher, M. A., Sprehe, M., Bartholomae, M., Hillen, W., and Brenn, R. G. (2011) Structures of carbon catabolite protein A–(HPr-Ser46-P) bound to diverse catabolite response element sites reveal the basis for high-affinity binding to degenerate DNA operators *Nucleic Acids Res.* 39, 2931–2942

34. Chen, Y. Q., Sengchanthalangsy, L. L., Hackett, A., and Ghosh, G. (2000) NF-κB (RelA) homodimer uses distinct mechanisms to recognize DNA targets. *Structure* 8, 419–428

35. Shimon, L. J. W., and Harrison, S. C. (1993) The phase 434 O2/R1–69 complex at 2.5 Å resolution. *J. Mol. Biol.* 232, 826–838

36. Rodgers, D. W., and Harrison, S. C. (1993) The complex between phase 434 repressor DNA-binding domain and operator site OR3: structural differences between consensus and non-consensus half-sites. *Structure* 1, 227–240

37. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M., and Harrison, S. C. (1988) Recognition of a DNA operator by the repressor of phase 434: a view at high resolution. *Science* 242, 899–907

38. Schwabe, J. W., Chapman, L., and Rhodes, D. (1995) The oestrogen receptor recognizes an imperfectly palindromic response element through an alternative side-chain conformation. *Structure* 3, 201–213