Crystal structure of the *Escherichia coli* Rob transcription factor in complex with DNA

Hyock Joo Kwon[1,2], Marjon H.J. Bennik[3,4], Bruce Demple[3] and Tom Ellenberger[1,2]

The *Escherichia coli* Rob protein is a transcription factor belonging to the AraC/XylS protein family that regulates genes involved in resistance to antibiotics, organic solvents and heavy metals. The genes encoding these proteins are activated by the homologous proteins MarA and SoxS, although the level of activation can vary for the different transcription factors. Here we report a 2.7 Å crystal structure of Rob in complex with the *micF* promoter that reveals an unusual mode of binding to DNA. The Rob–DNA complex differs from the previously reported structure of MarA bound to the *mar* promoter, in that only one of Rob's dual helix-turn-helix (HTH) motifs engages the major groove of the binding site. Biochemical studies show that sequence specific interactions involving only one of Rob's HTH motifs are sufficient for high affinity binding to DNA. The two different modes of DNA binding seen in crystal structures of Rob and MarA also match the distinctive patterns of DNA protection by AraC at several sites within the pBAD promoter. These and other findings suggest that gene activation by AraC/XylS transcription factors might involve two alternative modes of binding to DNA in different promoter contexts.

The *E. coli* Rob protein was first identified by its ability to bind to the right border of the *E. coli* origin of replication, *oriC*[1], although a role in DNA replication has not been identified. Rob is a member of the AraC/XylS family of transcription factors that regulate large and diverse groups of genes, termed regulons, in prokaryotes. Rob is most closely related to the SoxS, MarA, and TetD proteins[2]. The SoxS protein stimulates a group of genes whose products defend against oxidative damage (for example, superoxide dismutase) and it can mediate resistance to diverse antibiotics and organic solvents[3]. The MarA protein controls expression of some of the same genes as SoxS, including many that confer resistance to antibiotics[4]. The *marRAB* and the *soxRS* regulons include more than 15 genes each. Although the Rob protein's natural activity in *E. coli* is not fully understood, Rob can activate many of the same genes as SoxS and MarA[3–6]. Overexpression of Rob causes antibiotic resistance[6], organic solvent tolerance and heavy metal resistance[7], whereas deletion of Rob increases susceptibility to organic solvents[8].

Despite their broadly overlapping activities, MarA, SoxS, and Rob differentially activate the transcription of certain genes[2,5–12], suggesting that different types of interactions occur with the varied promoter sequences bound by these transcription factors. The SoxS and MarA proteins consist only of the conserved DNA binding domain (~100 residues) that is present in all AraC/XylS family members, whereas the Rob protein has an additional C-terminal domain (~200 residues) of unknown function. Overexpression of Rob's N-terminal DNA binding domain, corresponding to full length SoxS or MarA, confers multiple antibiotic resistance in *E . coli*[6], consistent with the idea that Rob's transcription activating function is housed within the conserved N-terminal domain.

Here we report the 2.7 Å crystal structure of the Rob transcription factor in complex with DNA containing the *micF* promoter sequence. The complex reveals two helix-turn-helix (HTH) motifs within Rob's N-terminal domain, an arrangement similar to the MarA protein[13], and a unique C-terminal domain that is structurally similar to the *E. coli* galactose-1-phosphate uridyltransferase enzyme (GalT). In the crystal structure, Rob inserts its N-terminal HTH into the major groove of an unbent DNA duplex. The C-terminal HTH motif lies on the surface of the DNA helix where it contacts the phosphodiester backbone. This mode of binding to DNA is consistent with biochemical studies of Rob's interactions with various promoter sequences and with chemical footprinting studies of AraC. In conjunction with the previously reported crystal structure of MarA[13], our results suggest that AraC/XylS transcription factors can engage promoter sites with high affinity in either of two alternative binding modes.

The structure of the Rob protein

The crystal structure of Rob bound to the *micF* promoter sequence (Fig. 1) was determined by multiple wavelength anomalous diffraction (MAD) methods using a selenomethionine substituted protein. The 2.7 Å structure of two Rob–DNA complexes within the crystallographic asymmetric unit has been refined to an R-factor of 25.4% (R_{free} = 30.2%; Table 1). In the crystals, two Rob monomers are associated with each *micF* DNA in two crystallographically independent complexes. One Rob protomer interacts in a specific manner with the *micF* DNA sequence (Fig. 2), contacting the major groove and DNA backbone over the consensus binding site. The other Rob protomer is bound nonspecifically on the opposite face of the DNA, one half turn (5–6 base pairs) away from the specific complex (Fig. 1*c*). This same arrangement of specific and nonspecific complexes is repeated in two crystallographically independent DNA complexes. Although the HTH motifs of the specific and nonspecific protomers are similarly positioned with respect to the major groove and DNA backbone, the electron density is poor or non-existent for the side chains facing the DNA in the nonspecific complex. In the DNA binding experiments described below, there is no evidence that multiple Rob protomers interact with a similar oligonucleotide in solution. These findings suggest that the end-to-end packing of DNA in the crystal lattice has created a nonspecific binding site for Rob that is only occupied

[1]Graduate Program in Biophysics, Harvard University, Cambridge, MA 02138, USA. [2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA. [3]Department of Cancer Cell Biology, Harvard School of Public Health, Boston, MA 02115, USA. [4]Present address: Agrotechnological Research Institute, Wageningen University Research Centre, 6700 AA Wageningen, The Netherlands.

Correspondence should be addressed to T.E. *email: tome@hms.harvard.edu*

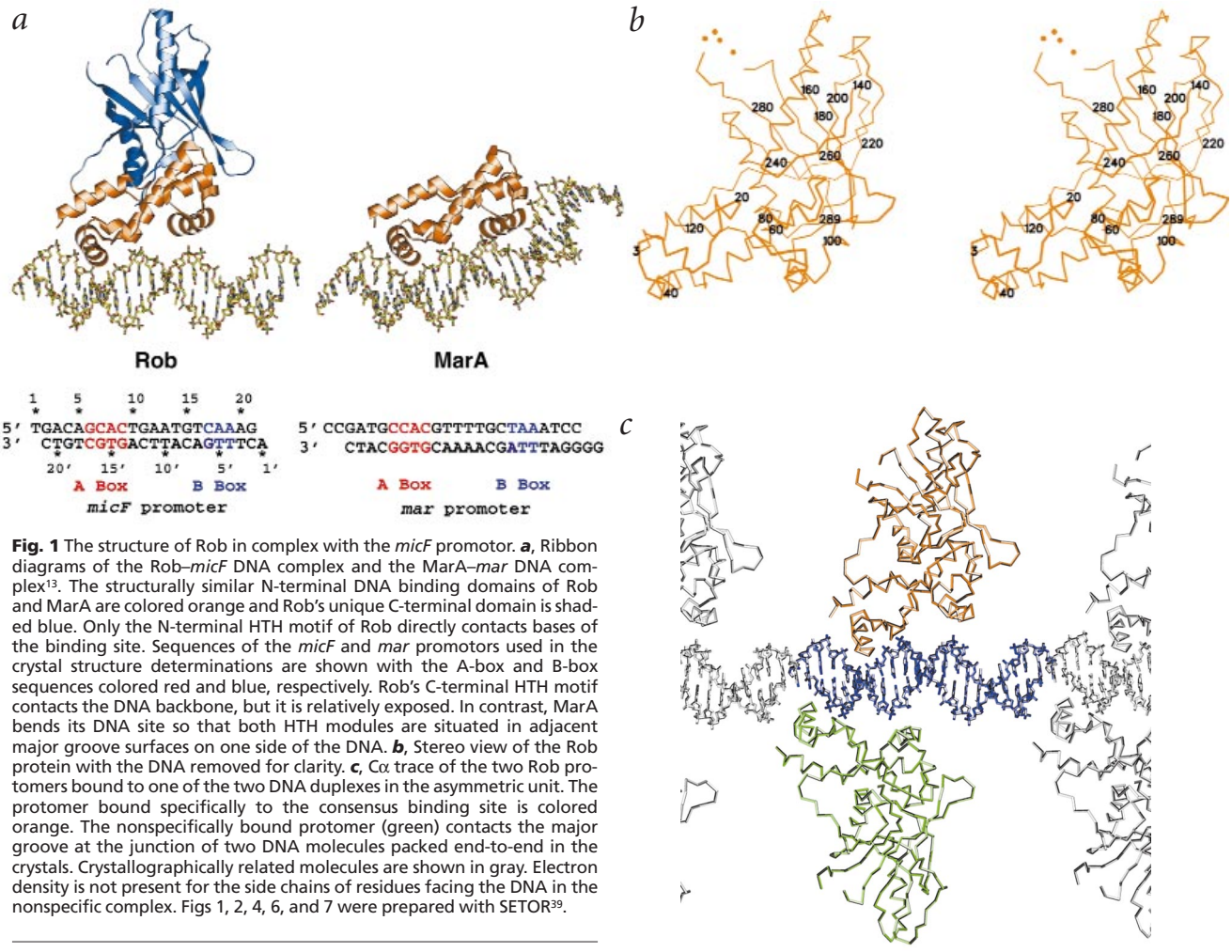


Fig. 1 The structure of Rob in complex with the *micF* promotor. ***a***, Ribbon diagrams of the Rob–*micF* DNA complex and the MarA–*mar* DNA complex[13]. The structurally similar N-terminal DNA binding domains of Rob and MarA are colored orange and Rob's unique C-terminal domain is shaded blue. Only the N-terminal HTH motif of Rob directly contacts bases of the binding site. Sequences of the *micF* and *mar* promotors used in the crystal structure determinations are shown with the A-box and B-box sequences colored red and blue, respectively. Rob's C-terminal HTH motif contacts the DNA backbone, but it is relatively exposed. In contrast, MarA bends its DNA site so that both HTH modules are situated in adjacent major groove surfaces on one side of the DNA. ***b***, Stereo view of the Rob protein with the DNA removed for clarity. ***c***, Cα trace of the two Rob protomers bound to one of the two DNA duplexes in the asymmetric unit. The protomer bound specifically to the consensus binding site is colored orange. The nonspecifically bound protomer (green) contacts the major groove at the junction of two DNA molecules packed end-to-end in the crystals. Crystallographically related molecules are shown in gray. Electron density is not present for the side chains of residues facing the DNA in the nonspecific complex. Figs 1, 2, 4, 6, and 7 were prepared with SETOR[39].

at the high protein concentrations used for crystal growth. The specific interactions of the Rob protein with the *micF* promoter sequence are described in detail below.

The Rob protein consists of an N-terminal DNA binding domain that is homologous to the MarA protein[13], and a unique C-terminal domain of unknown function (Fig. 1). The N-terminal DNA binding domain contains two HTH motifs connected by a rigid central helix that fixes their relative orientations. The C-terminal domain is composed of an eight-stranded antiparallel β-sheet, sandwiched between a pair of α-helices and a two-stranded antiparallel β-sheet (Fig. 1). The C-terminal domain sits on top of the N-terminal domain, capping the C-terminal HTH motif and burying more than 1,200 Å^2 of its surface area. The interface between these domains consists of hydrophobic residues from the N-terminal (Val 64, Leu 68, and Trp 109) and the C-terminal (Leu 149, Ile 152, and Tyr 244) domains, as well as an electrostatic interaction between Glu 108 and Arg 288.

DNA recognition by Rob

Rob and the related transcription factors SoxS and MarA bind to asymmetric DNA sequences ~20 base pairs in length within the promoter regions of regulated genes. These binding sites are composed of two conserved sequence motifs, termed the A-box and B-box, separated by seven base pairs of variable sequence (Fig. 1*a*). Although both HTH motifs within Rob's DNA binding domain have the potential to interact with DNA, only the N-terminal HTH reading head closely engages the DNA base pairs in the Rob crystal structure (Fig. 1*a*). Rob's N-terminal HTH motif lies in the major groove over the A-box, making both van der Waals and hydrogen bond interactions with the base pairs of the A-box sequence (Fig. 2*a*). The C-terminal HTH motif does not lie in the major groove of the B-box, but instead contacts the DNA backbone (Fig. 2*b*). This mode of binding to DNA is seen for two independent Rob–DNA complexes within the crystal asymmetric unit.

Base specific contacts to the A-box sequence uniquely specify Rob's interaction with this half of the binding site. Residues in the second helix (αC) of the N-terminal HTH motif complement the shape and electrostatic properties of the major groove of the A-box sequence, spanning nucleotides 6–9 and 14′–17′ of the DNA (Figs 1*a*, 2*a*, 3). Trp 36 makes van der Waals contact with Cyt 7, whereas Arg 40 is hydrogen bonded to the O6 of Gua 6. The Nε nitrogen of Trp 36 is 3.2 Å from the O4 of Thy 15′, although the geometry is not ideal for a hydrogen bond. Gln 39 is within 3.3 Å of Gua 14′, in a position similar to the analogous glutamine of MarA in the MarA–DNA structure[13]. The interactions between Rob and the A-box sequence are very analogous to those described for the crystal structure of MarA in complex with the *mar* promoter site. Unlike MarA, Rob does not utilize its C-terminal HTH motif to interact sequence specifically with the major groove of the B-box in the crystal structure (Fig. 1*a*).

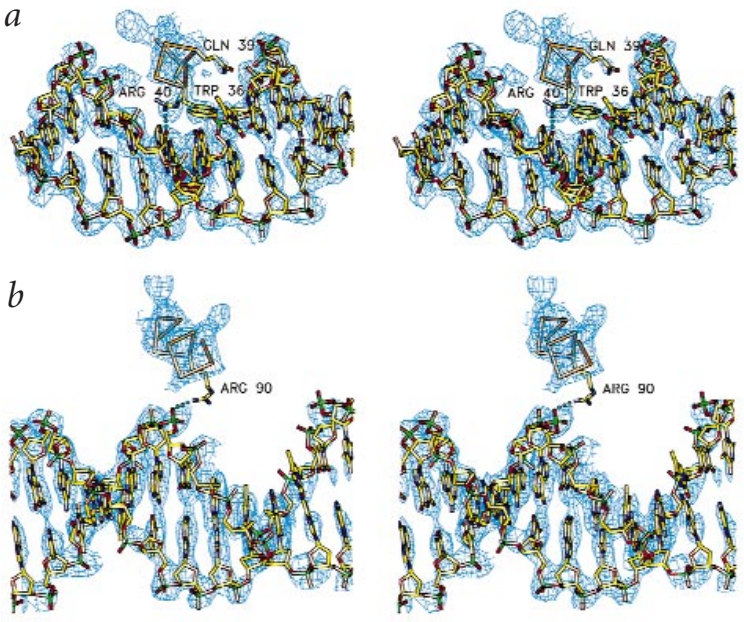


Fig. 2 Interactions of Rob with DNA. ***a***, A close-up view of the N-terminal HTH reading head (helix αC) of Rob and its interactions with the A-box of the *micF* binding site. The experimentally phased electron density after solvent flattening is shown at a contour level of 1σ above the mean value in both panels. Arg 40 forms a hydrogen bond with a conserved base Gua 6 (compare Fig. 1*a*), while Trp 36 and Gln 39 are involved in van der Waals interactions with the DNA. ***b***, Close-up of the C-terminal HTH reading helix (αF) of Rob and its interactions with the B-box. The conserved residue Arg 90 contacts the phosphate backbone. The side chain of Arg 90 is clearly defined in the final $2F_o$ - F_c electron density (not shown). In the MarA–DNA complex (Fig. 1*a*) the homologous residue Arg 96 forms hydrogen bonds with two guanines of the DNA binding site.

Rob and MarA engage DNA differently

The N-terminal DNA binding domain of Rob is clearly homologous to the full length MarA protein (51% sequence identity; Fig. 3) and its structure is also conserved[13] (Fig. 1*a*). The tandem HTH motifs of Rob and MarA exhibit a root mean square (r.m.s.) deviation of 0.9 Å, calculated for all main chain atoms within the conserved domain. Despite this close structural similarity, Rob contacts the *micF* promoter differently from MarA's interaction with the *mar* promoter. In two independent protein–DNA complexes helix αC from the N-terminal HTH motif makes similar contacts with the major groove of the A-box DNA sequence (described above). However, there is a striking difference in the interaction of the C-terminal HTH motif with DNA in the Rob and MarA complexes (Figs 1*a*, 4). In MarA, both HTH motifs are positioned in the major groove of the *mar* DNA, requiring a 35° bend in the DNA to establish a close fit between the protein and edges of base pairs of both the A-box and B-box. Arg 96, located on helix αF of MarA's C-terminal HTH motif, forms direct hydrogen bonds with two guanines (corresponding to Gua 15 and Ade 7′ of *micF*; Fig. 1*a*) in the major groove adjacent to the B-box sequence. In principle, these hydrogen bonds could contribute to sequence specific binding by MarA, but the guanines contacted by Arg 96 are not conserved in the binding sites of Rob, MarA, and SoxS[12] (Fig. 1*a*). The remaining contacts between MarA's C-terminal HTH motif and the B-box of the *mar* promoter include water mediated hydrogen bonds to nonconserved bases and van der Waals interactions[13]. These interactions are unlikely to make a major contribution to binding specificity and they are apparently not required for high affinity interactions with the *mar* and *micF* DNA sites (see below). In the structure of Rob in complex with the *micF* promoter, the DNA is straight and in the B-form (Fig. 1*a*), and the protein does not directly contact the bases within and around the B-box sequence. Instead, Arg 90 from the second helix (αF) of the C-terminal HTH motif forms a hydrogen bond with the phosphate of Thy 14 (Fig. 2*b*). This nonspecific interaction of Arg 90 with the DNA backbone presumably compensates for the lack of major groove interactions with the B-box and it contributes to Rob's affinity for the *micF* promoter, which is comparable to MarA's affinity for its binding site (Fig. 5).

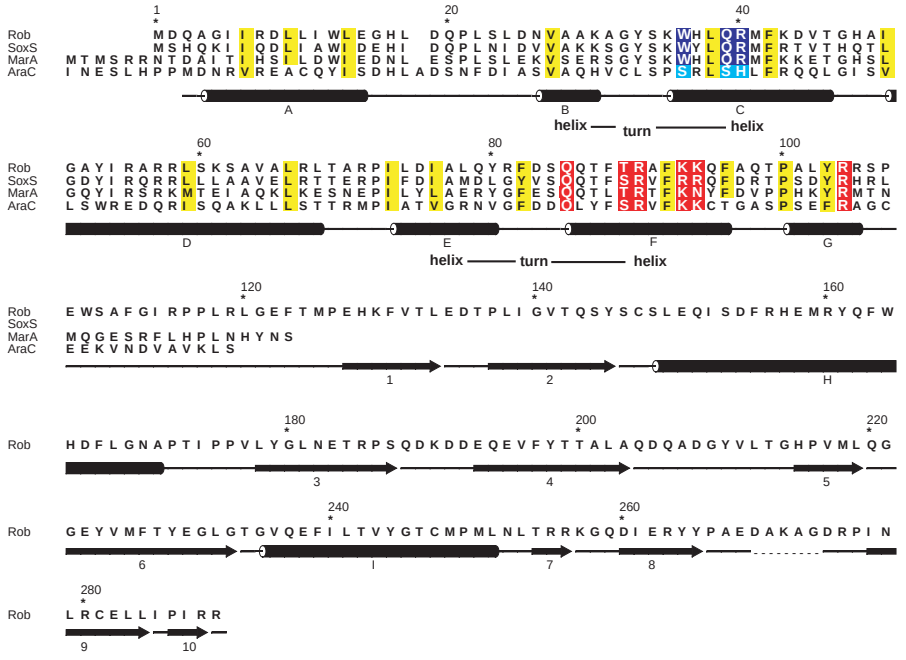


Fig. 3 Sequence alignment of AraC/XylS family members. The aligned sequences of Rob, MarA, and AraC are shown with secondary structure (α-helices as tubes, β-strands as arrows) of the Rob protein displayed below the alignment. The residue numbers correspond to the Rob protein sequence. Conserved, buried hydrophobic residues in the dual HTH DNA binding motif are highlighted in yellow. Residues in the N-terminal HTH motif of Rob and MarA that contact conserved bases in the major groove of the A-box are highlighted in blue. The AraC residues at these positions (cyan) are not conserved. A highly conserved cluster of polar residues in the C-terminal HTH motif (shaded red) contact the bases and phosphodiester backbone of the B-box in the MarA–DNA complex[13]. These residues are solvent exposed in the Rob–DNA complex. Residues within the adjacent helix αG are also highly conserved[2]. This figure was prepared with ALSCRIPT[40].

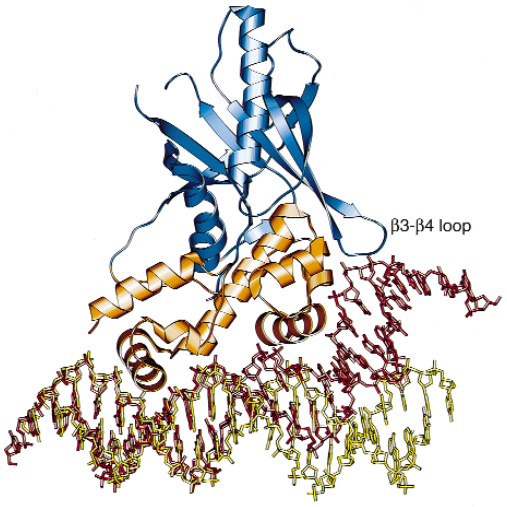


Fig. 4 Rob's C-terminal domain might restrict DNA bending. The structure of Rob is shown in complex with DNA from the crystal structure (yellow), and with the DNA from the MarA crystal structure[13] (red) superimposed on the Rob protein. The bent DNA from the MarA complex clashes with the acidic β3-β4 loop (compare Fig. 3) that projects from the C-terminal domain of Rob (right side of diagram). The loop and the β-strands on either side of the loop would have to rearrange to accommodate the modeled interaction of Rob with bent DNA.

A loop connecting β-strands 3 and 4 (residues 187–193) in Rob's C-terminal domain is located near the DNA binding surface where it might prevent DNA from bending toward the protein. The β3-β4 loop is composed primarily of acidic residues (Fig. 3) that would be in close contact with DNA if Rob were to bend the DNA in a manner similar to the MarA protein (Fig. 4). MarA lacks an analogous loop, allowing the DNA to wrap around MarA's compact protein fold without interference. A rearrangement of the loop might permit the close approach of a bent DNA to Rob. However, a loop-deleted mutant of Rob (deletion of residues 187–193) binds to the 21 base pair *micF* DNA fragment with the same affinity as wild type Rob protein (data not shown). This finding suggests either that the acidic loop is highly distensible and it does not interfere with DNA bending, or that Rob prefers to bind to a straight DNA (Fig. 1*a*) irrespective of the presence or absence of the acidic loop. Rob was independently identified in a screen for proteins that bind to pre-bent DNA sequences[14]. However, the sequence of the bent DNA (gGCAaaaacgggCAAaaacg, where the upper case letters are conserved bases of the A- and B-boxes) used in this study contains a nearly perfect consensus binding site for Rob. Thus, possible effects of DNA bending on binding site selection by the Rob are obscured by the sequence of the binding site used in this study. Clearly, Rob can bind to bent DNA and has been shown to bend both the *fumC* and *zwf* sequences in the context of longer DNA molecules[9]. However, the results presented below show that close packing interactions between Rob and the major groove surface of the B-box are not required for high affinity binding to DNA.

The C-terminal domain of Rob

Of the 100 or more transcription factors belonging to the AraC/XylS family, only seven are similar to MarA in that they consist solely of a DNA binding domain[2]. The other members of this large family contain additional nonconserved domains, which in all but two cases are of unknown function. The well characterized transcription factor AraC contains an N-terminal arabinose binding domain that regulates the dimerization of AraC and thus transcription[15]. XylS also contains a nonconserved domain that is suspected of interacting with the effector molecule benzoate. XylS is postulated to exist in equilibrium between transcriptionally active and inactive forms[16]. It is therefore tempting to speculate that some effector molecule might bind to the C-terminal domain of the Rob protein, which is expressed at relatively high levels in growing cells[1].

An automated search of the Protein Data Bank[17] revealed that the C-terminal domain of Rob is structurally homologous to a portion of *E. coli* GalT[18]. GalT converts galactose-1-phosphate and UDP-glucose to glucose-1-phosphate and UDP-galactose. The structure of the C-terminal domain of Rob (residues 121–289) superimposes well on residues 125–176 and 234–304 of GalT, including the UDP-galactose/UDP-glucose binding pocket (Fig. 6). GalT cradles UDP-galactose between an antiparallel β-sheet and its connecting loop. A second N-terminal loop of GalT also helps form the pocket for UDP-galactose, but this region of GalT is not homologous to any portion of the Rob C-terminal domain. The roof of the UDP-galactose binding pocket of GalT is provided by the opposite subunit of the GalT dimer. Rob lacks an analogous structure. The binding pocket for UDP-galactose is composed almost entirely of hydrophilic residues that contribute a number of hydrogen bonds

a

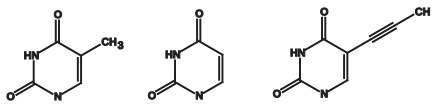
K_{DNA}(nM)

	micF	micFU	micFP	mar	marU	marP
Rob	6.0	6.0	5.4	4.4	5.7	5.5
MarA	41.0	50.5	78.0	43.0	61.5	72.0

b

```
micF     5′  GTATTTGACAGCACTGAATGTCAAAACAAAACCTT
         3′  CATAAACTGTCGTGACTTACAGTTTTGTTTTGGAA

micFU    5′  GTATTTGACAGCACTGAATGTCAAAACAAAACCTT
         3′  CATAAACTGTCGTGACTTACAGUUTTGTTTTGGAA

micFP    5′  GTATTTGACAGCACTGAATGTCAAAACAAAACCTT
         3′  CATAAACTGTCGTGACTTACAGPTTTGTTTTGGAA

mar      5′  TTGACCGATGCCACGTTTTGCTAAATCGGTTCAAG
         3′  AACTGGCTACGGTGCAAAACGATTTAGCCAAGTTC

marU     5′  TTGACCGATGCCACGTTTTGCTAAATCGGTTCAAG
         3′  AACTGGCTACGGTGCAAAACGAUUTAGCCAAGTTC

marP     5′  TTGACCGATGCCACGTTTTGCTAAATCGGTTCAAG
         3′  AACTGGCTACGGTGCAAAACGAPTTAGCCAAGTTC
                        A Box       B Box
```

c



T : Thymine
U : Uracil
P : 5-(1-propynyl)-Uracil

Fig. 5 DNA binding affinities for modified promoter sequences. ***a***, The apparent DNA binding affinities (K_{DNA}) of the Rob and MarA proteins were determined for the wild type *micF* and *mar* promoter sites, and for sites with the modified B-box sequences shown in ***b***. ***c***, The thymine (T) analogs uracil (U) and 5-(1-propynyl)-uracil (P) were incorporated at the indicated positions within the B-box sequence, and DNA binding affinities were determined by an electrophoretic mobility shift assay. Removal of the thymine methyl group by uracil substitutions or the addition of a bulky propynyl group in the major groove has little effect on the binding affinities of Rob and MarA for these DNA sites. These results imply that close interactions with the major groove of the B-box are not required for high affinity binding to the *micF* and *mar* promoters.

Fig. 6 Structural similarity of Rob and GalT. Superposition of the C-terminal domain of Rob with the UDP-galactose binding pocket of GalT[18]. UDP-galactose is cradled by an antiparallel β-sheet (green) and the connecting loop (purple) in the GalT structure. The analogous antiparallel sheet from Rob (yellow) superimposes well, except that the acidic loop (red) is extended in comparison to the analogous loop of GalT (purple). The portion of the Rob C-terminal domain that has no counterpart in GalT is shown in blue.

to UDP-galactose. Superposition of the two structures suggests a plausible ligand binding pocket on the surface of Rob. This corresponding pocket in Rob includes an antiparallel β-strand and its connecting loop (Figs 4, 6), which is the acidic β3-β4 loop spanning residues 187–193 (described above) that has an extended conformation in the Rob–DNA structure.

The structural similarity of Rob's C-terminal domain and the GalT ligand binding pocket suggests a model in which the C-terminal domain of Rob might bind to an effector molecule that regulates the transcriptional activity of Rob. The bound ligand could induce Rob's acidic loop to curl away from the DNA binding surface, adopting a conformation similar to that of the analogous loop of GalT in complex with UDP-galactose (Fig. 6). However, significant differences between the ligand binding residues of GalT and the superimposed residues of Rob do not immediately suggest candidate ligands for the Rob protein. Equilibrium dialysis binding experiments using radiolabeled UDP-galactose, UDP-glucose, or UDP-galactosamine did not reveal a ligand interaction with the Rob protein in complex with the *micF* promoter DNA (data not shown). Recent genetic data show that Rob represses transcription of the *galT* gene (M.H.J.B. and B.D., unpublished results). It is possible that some other intermediate of galactose metabolism could be an effector molecule for Rob.

The full length MarA protein and Rob's analogous N-terminal domain both function as transcriptional activators[6,9], implying that similar features of these structurally homologous proteins are directly involved in gene activation. The related transcription factors SoxS and AraC have been shown to enhance the binding of RNA polymerase to DNA[19,20], and there is evidence for direct interactions of Rob and SoxS with the α-subunit of RNA polymerase that could assist in the initiation of transcription[9,21]. In the Rob–DNA complex, most of the surface of the conserved N-terminal domain facing towards the transcriptional start site is either buried in the protein–DNA interface or packed against the C-terminal domain. In contrast, a portion of the C-terminal HTH motif is more exposed. The acidic β3-β4 loop is situated on the protein surface facing towards the transcription start site of the *micF* promoter, where it obscures most of helix αG except for its C-terminal end, which is exposed in the crystal structure (Figs 1*a*, 4). The bent DNA conformation seen in the MarA–DNA complex would further sequester this surface of the Rob protein in the interface with DNA.

The sequences of more than 100 members of the AraC/XylS family of transcription factors contain a cluster of highly conserved polar residues in the region of the C-terminal HTH motifs of proteins that bind to very different DNA sequences[2] (Fig. 3). Two of these conserved residues (Arg 90, Lys 94) contact the DNA backbone in the Rob–DNA complex. The others decorate the exposed surface of the C-terminal HTH where they are available for interactions with other molecules. This conserved region might have a function separate from DNA recognition, such as the activation of transcription, perhaps by interacting with RNA polymerase.

The importance of the B-box

Electrophoretic gel mobility shift assays using modified DNA substrates were performed with both Rob and MarA in order to assess the energetic importance of protein interactions with the B-box (Fig. 5). Two conserved thymines within the B-box (positions 4′ and 5′) have been suggested to contribute to sequence specific interactions with MarA in the form of van der Waals contacts to the C5 methyl groups[13]. We replaced both of these thymines with uracil, thereby eliminating contacts to the C5 methyl groups, and determined the binding affinity of MarA and Rob for the modified *mar* site, as well as the affinity for a similarly modified *micF* promoter site. Rob and MarA bind to the uracil containing promoter sequences with apparent affinities that are indistinguishable from those obtained for the unmodified promoter site (Fig. 5). In order to assess whether helix αF must come in close contact with the B-box, we replaced the thymine C5 methyl groups at these positions of the DNA with a larger propyne group (5-(1-propynyl)-uracil; Fig. 5). This bulky substitution is located in the major groove of the B-box and it would hinder the C-terminal HTH from inserting into the major groove of the bent DNA as seen in the MarA–DNA crystal structure. Both Rob and MarA bind the propyne containing promoter sequence with affinities that are similar to those of the unmodified promoters, although MarA binding was affected somewhat more than Rob binding (Fig. 5). These results

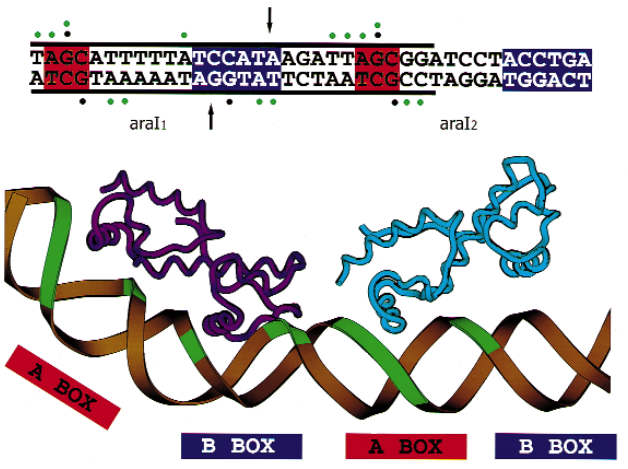


Fig. 7 A model of AraC in complex with the pBAD promoter. The published chemical footprinting data for AraC in complex with the $araI_1$ and $araI_2$ sites of the pBAD promoter[27] are summarized. The positions of methylated guanines (black dots) and ethylated phosphates (green dots) that interfere with AraC binding are indicated. The arrows indicate DNase I hypersensitive sites in the AraC–DNA complex, and the black horizontal bar shows the region protected from DNase I digestion. These footprinting data can be accounted for by the theoretical model of the AraC–DNA complex shown in the bottom half of the figure. The structures of the MarA–*mar* DNA complex[13] (purple) and the Rob–*micF* DNA complex (cyan) best fit the chemical protection pattern observed for the $araI_1$ and $araI_2$ sites, respectively. The positions of phosphates shown by ethylation interference to interact with AraC[27] are highlighted in green on the DNA model.

Table 1 Data collection and refinement statistics for the Rob–DNA complex

	λ1 Se-Met	λ2 Se-Met	λ3 Se-Met	Native
Wavelength (Å)	0.9783	0.9789	0.9500	0.9161
Resolution (Å)	2.8	2.8	2.8	2.7
Reflections				
Total	413,049	310,353	230,194	306,889
Unique	51,976	51,367	52,182	54,476
Completeness (%)	97.8	98.8	97.3	95.1
I / σ(I)	8.1	10.1	9.5	8.7
R_{sym}[1]	9.6	7.7	7.5	8.2
R_{cullis}[2] (anomalous)	0.76 (0.80)	0.72 (0.90)	(0.90)	
Phasing power[3] (centric/acentric)	1.04 / 0.78	1.2 / 0.97	– / –	
Figure of merit			0.47	
Refinement				
Resolution (Å)	30.0–2.7			
R_{cryst}/R_{free}[4]	25.4% / 30.2%			
R.m.s. deviation				
Bonds (Å)	0.0075			
Angles (°)	1.3604			

[1] $R_{sym} = \Sigma|I_j - \langle I\rangle| / \Sigma\langle I\rangle$, where I_j is the intensity of reflection j and $\langle I\rangle$ is the mean intensity.
[2] $R_{cullis} = \Sigma||F_{ph} \pm F_p| - F_{h,c}| / \Sigma|F_{ph} + F_p|$, where $F_{h,c}$ is the calculated heavy atom structure factor.
[3] Phasing power = $\langle F_h\rangle / E$, where $\langle F_h\rangle$ is the root mean square heavy atom structure factor and E is the residual lack of closure.
[4] $R_{cryst} = \Sigma||F_o| - |F_c|| / \Sigma|F_o|$, calculated with the working reflection set. R_{free} is the same as R_{cryst} but calculated with the reserved reflection set.

show that the close packing of the C-terminal HTH motif of Rob (and perhaps MarA) with the bases of the B-box sequence is not necessary for high affinity interactions with DNA, and that the B-box sequence makes an apparently modest contribution to the promoter specific binding. Rob binds to the 20-mer duplex used for crystallization (data not shown) with an apparent affinity that is indistinguishable from Rob's affinity for the longer 35-mer duplex reported in Fig. 5, making it unlikely that the DNA conformation in the Rob crystal structure results from a DNA binding site that is too short. *In vitro* binding of Rob to the *micF* promoter[6,22], and transcriptional activation of *micF in vitro*[9] have been demonstrated. It was shown by northern analysis that the basal expression of *micF* is significantly higher in wild type *E. coli* than in a *rob* deleted mutant, establishing the relevance of Rob dependent expression of *micF in vivo* (M.H.J.B. and B.D, unpublished results). The same conclusion was reached in an independent study[23].

The foregoing conclusions are consistent with the reported mutational analysis of the *micF* promoter[22]. Single base pair changes within the A-box of the *micF* promoter decreased Rob's binding affinity 7–100-fold. Sequence changes affecting the B-box of the *micF* promoter, however, had only a 2–4-fold effect on Rob's DNA binding affinity[22] (Fig. 5). Likewise, the promoters of genes regulated by MarA and SoxS contain putative binding sites with diverse B-box sequences that are not suggestive of sequence specific interactions between the B-box and these proteins. These results strongly suggest that most of the sequence specific DNA interactions made by Rob and related transcription factors involve the A-box of the binding site (consistent with the crystal structure of the Rob–DNA complex), although B-box interactions might be important for enhanced discrimination of different promoter sites.

Dual modes of binding

A mutational analysis of binding sites for the *E. coli* AraC protein has identified protein contacts with two separate regions of a 17 base pair consensus binding site[24] that are analogous to the A-box and B-box of the *micF* promoter. Within the pBAD promoter, two AraC binding sites ($araI_1$ and $araI_2$) are arranged in a direct repeat (Fig. 7) whose spacing and orientation are important for transcriptional activation[25]. Mutational data show that AraC's N-terminal HTH motif contacts the A-box of $araI_1$[26], whereas the B-box is contacted by the C-terminal HTH motif[24]. The $araI_2$ site, which is located proximal to the transcription start site, lacks an optimal B-box[24]. Chemical footprinting data reveal little or no interaction with the B-box of $araI_2$[27]. A model can be constructed for the interaction of AraC with the pBAD promoter that is consistent with the DNA footprinting data and both modes of DNA recognition seen in structures of the Rob and MarA transcriptional activators bound to their DNA sites. In the model, one AraC protomer binds to the A-box and B-box of a bent $araI_1$ DNA site (similar to the MarA–DNA crystal structure) and a second AraC protomer makes specific interactions with only the A-box of the $araI_2$ site (the situation seen in our Rob–DNA crystal structure; Fig. 7). Although the B-box region of the $araI_2$ site is important for transcriptional activation[25], the footprinting data for AraC show little or no interaction with this half of the $araI_2$ binding site.

In nearly all the promoter binding sites of AraC/XylS family members that have been mapped by DNase I footprinting, the B-box is proximal to the transcriptional start site. Exceptions, however, include the *mar*, *soxS*, and *ribA* promoters that contain functional binding sites in which the A-box is proximal to the transcriptional start site[12,23,28,29]. Inverted sites can stimulate transcription[12,23,28,29] and might be inhibitory in some cases[30]. The DNA site bound by MarA in the crystal structure[13] is in the 'reverse' orientation within the *mar* promoter[12]. This *mar* promoter site will support activation of a reporter gene by MarA or SoxS when it is substituted for the native binding site of the *micF* promoter, but only if the substituted *mar* sequence is placed in the forward orientation, the opposite of its orientation in the *mar* promoter[12]. Similar conclusions have been reached for SoxS dependent promoters[31]. These findings show that Rob, MarA, and SoxS function in two different orientations in different promoter contexts. The two different modes of interaction with DNA seen in the crystal structures of Rob and MarA could relate to different requirements within different promoter contexts and/or different steps in the activation of gene expression.

Methods

Purification and crystallization of the Rob protein. The coding sequence for the Rob protein was cloned between the *Nco*I and *Hin*dIII sites of the pET28b vector (Novagen) resulting in a C-terminal extension of 14 residues containing a hexahistidine affinity tag. The histidine-tagged Rob protein was expressed in *E. coli* and purified by nickel-chelate affinity chromatography. Pure fractions were pooled, concentrated, and subjected to gel filtration chromatography. Following gel filtration, pure fractions were pooled and concentrated in 20 mM Tris pH 8.0, 200 mM NaCl, 10% (v/v) glycerol, and 1 mM DTT to a final concentration of 19 mg ml^{-1}. Selenomethionyl (Se-Met) Rob was prepared in *E. coli*[32] and purified under strictly reducing conditions. A series of synthetic DNA oligonucleotides containing the *micF* promoter sequence were synthesized by standard phosphoramidite chemistry. Oligonucleotides were purified by denaturing gel electrophoresis. Annealed duplexes were mixed with Rob in a 1:2 molar ratio of protein to DNA at a final protein concentration of 5 mg ml^{-1}. Crystals were grown by mixing equal volumes of the Rob–DNA complex with a reservoir solution consisting of 100 mM MES pH 6.5, 125 mM $MgCl_2$, 10% (v/v) glycerol, 1 mM DTT, 1 mM spermine-HCl, and 15% (w/v) PEG 8000.

X-ray data collection and structure determination. Crystals were transferred into a solution containing 100 mM MES pH 6.5, 125 mM $MgCl_2$, 20% (v/v) glycerol, and 20% (w/v) PEG 8000, and soaked for 5–12 h. The soaking procedure greatly reduced the mosaicity of the crystals and resulted in higher quality X-ray data. Crystals were flash-frozen in a nitrogen stream at -160 °C and stored in liquid nitrogen. Diffraction data were processed with DENZO and SCALEPACK[33]. Crystals belong to space group $P2_1$ with unit cell dimensions: a = 75.3 Å, b = 206.7 Å, c = 67.4 Å, $\alpha = \beta = \gamma = 90.0°$. The asymmetric unit of the crystal contains four Rob molecules and two 21mer duplex oligonucleotides.

MAD data (Table 1) were collected from crystals of the Se-Met labeled protein at beamline X-25 of the National Synchrotron Light Source (Upton, NY). The Se positions were located in difference Patterson maps using the differences between λ2 (the selenium K-edge) and λ3 (the remote high energy wavelength). Nine selenium sites out of 32 possible sites (four protomers with eight methionines each) were found using the program RSPS[34] and an additional seventeen sites were found in difference Fourier maps that were phased using the original nine Se sites. Heavy atom parameters were refined with the program MLPHARE[34] and the subsequent electron density maps were improved by solvent flipping with the program SOLOMON[35]. The resulting electron density was readily interpretable (Fig. 2) and the protein chain was built guided by the positions of the Se sites. The register of the DNA sequence was confirmed by using the positions of iodine sites from a data set collected from a crystal containing 5-iodouracil-substituted DNA (data not shown). Several cycles of refinement with CNS[36] using torsion angle molecular dynamics and manual rebuilding of the model with O[37] resulted in a model with an R-factor of 25.4% (R_{free} = 30.2%)[38] using all data in the resolution range 30.0–2.7 Å. The high value of the Wilson B-factor for the X-ray intensity data (70 Å^2) contributes to the high average B-factor of the model.

DNA binding assays. Phosphoramidites for uracil and 5-(1-propynyl)-uracil were purchased from Glen Research and incorporated into a 35mer oligonucleotide corresponding to either the *micF* or *mar* promoter sequence (Fig. 5) by standard automated synthesis. The oligonucleotides were purified by denaturing gel electrophoresis and labeled at their 5´ ends with T4 polynucleotide kinase and [γ-^{32}P]-ATP. Various amounts of Rob or MarA protein were mixed with 0.01 pmol of DNA (0.5 nM final concentration) in 20 μl of binding buffer (20 mM Tris pH 8.0, 50 mM NaCl, 5% glycerol, 0.5% (v/v) NP40). The binding reactions were incubated at room temperature for 15 min then loaded onto 10% polyacrylamide gels and electrophoresed at 250 V at 4 °C. The amounts of bound and free DNA were analyzed by autoradiography.

Coordinates. Coordinates of the Rob/*mic*F complex have been deposited in the Protein Data Bank (accession code 1D5Y).

Acknowledgments

We thank A. Lau, M. Sawaya and members of the Hogle research group for assistance with X-ray data collection. We appreciate the generous gift of purified MarA protein from M. Alekshun and S. Levy (Tufts University School of Medicine). This work was supported by grants from the National Institutes of Health (T.E, B.D.), a Howard Hughes Medical Institute Predoctoral Fellowship (H.J.K.), and the Agrotechnological Research Institute, The Netherlands (M.H.J.B.). We also acknowledge the support of the Giovanni Armenise Foundation for Advanced Scientific Research and the Harvard Center for Structural Biology.

Received 28 December, 1999; accepted 29 March, 2000.

1. Skarstad, K., Thony, B., Hwang, D.S. & Kornberg, A. A novel binding protein of the origin of the *Escherichia coli* chromosome. *J. Biol. Chem.* **268**, 5365–5370 (1993).
2. Gallegos, M.T., Schleif, R., Bairoch, A., Hofmann, K. & Ramos, J.L. Arac/XylS family of transcriptional regulators. *Microbiol. Mol. Biol. Rev.* **61**, 393–410 (1997).
3. Demple, B. Redox signaling and gene control in the *Escherichia coli soxRS* oxidative stress regulon — a review. *Gene* **179**, 53–57 (1996).
4. Alekshun, M.N. & Levy, S.B. Regulation of chromosomally mediated multiple antibiotic resistance: the mar regulon. *Antimicrob. Agents. Chemother.* **41**, 2067–2075 (1997).
5. Cohen, S.P., Hachler, H. & Levy, S.B. Genetic and functional analysis of the multiple antibiotic resistance (mar) locus in *Escherichia coli*. *J. Bacteriol.* **175**, 1484–1492 (1993).
6. Ariza, R.R., Li, Z., Ringstad, N. & Demple, B. Activation of multiple antibiotic resistance and binding of stress-inducible promoters by *Escherichia coli* Rob protein. *J. Bacteriol.* **177**, 1655–1661 (1995).
7. Nakajima, H., Kobayashi, K., Kobayashi, M., Asako, H. & Aono, R. Overexpression of the *robA* gene increases organic solvent tolerance and multiple antibiotic and heavy metal ion resistance in *Escherichia coli*. *Appl. Environ. Microbiol.* **61**, 2302–2307 (1995).
8. White, D.G., Goldman, J.D., Demple, B. & Levy, S.B. Role of the *acrAB* locus in organic solvent tolerance mediated by expression of marA, soxS, or robA in *Escherichia coli*. *J. Bacteriol.* **179**, 6122–6126 (1997).
9. Jair, K.W. *et al.* Transcriptional activation of promoters of the superoxide and multiple antibiotic resistance regulons by Rob, a binding protein of the *Escherichia coli* origin of chromosomal replication. *J. Bacteriol.* **178**, 2507–2513 (1996).
10. Greenberg, J.T., Chou, J.H., Monach, P.A. & Demple, B. Activation of oxidative stress genes by mutations at the *soxQ/cfxB/marA* locus of *Escherichia coli*. *J. Bacteriol.* **173**, 4433–4439 (1991).
11. Ariza, R.R., Cohen, S.P., Bachhawat, N., Levy, S.B. & Demple, B. Repressor mutations in the *marRAB* operon that activate oxidative stress genes and multiple antibiotic resistance in *Escherichia coli*. *J. Bacteriol.* **176**, 143–148 (1994).
12. Martin, R.G., Gillette, W.K., Rhee, S. & Rosner, J.L. Structural requirements for marbox function in transcriptional activation of *mar/sox/rob* regulon promoters in *Escherichia coli*: sequence, orientation and spatial relationship to the core promoter [In Process Citation]. *Mol. Microbiol.* **34**, 431–441 (1999).
13. Rhee, S., Martin, R.G., Rosner, J.L. & Davies, D.R. A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. *Proc. Natl Acad. Sci. USA* **95**, 10413–10418 (1998).
14. Kakeda, M., Ueguchi, C., Yamada, H. & Mizuno, T. An *Escherichia coli* curved DNA-binding protein whose expression is affected by the stationary phase-specific sigma factor sigma S. *Mol. Gen. Genet.* **248**, 629–634 (1995).
15. Lobell, R.B. & Schleif, R.F. DNA looping and unlooping by AraC protein. *Science* **250**, 528–532 (1990).
16. Ramos, J.L., Michan, C., Rojo, F., Dwyer, D. & Timmis, K. Signal-regulator interactions. Genetic analysis of the effector binding site of xylS, the benzoate-activated positive regulator of Pseudomonas TOL plasmid meta-cleavage pathway operon. *J. Mol. Biol.* **211**, 373–382 (1990).
17. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
18. Thoden, J.B., Ruzicka, F.J., Frey, P.A., Rayment, I. & Holden, H.M. Structural analysis of the H166G site-directed mutant of galactose-1-phosphate uridylyltransferase complexed with either UDP-glucose or UDP-galactose: detailed description of the nucleotide sugar binding site. *Biochemistry* **36**, 1212–1222 (1997).
19. Li, Z. & Demple, B. SoxS, an activator of superoxide stress genes in *Escherichia coli*. Purification and interaction with DNA. *J. Biol. Chem.* **269**, 18371–18377 (1994).
20. Zhang, X., Reeder, T. & Schleif, R. Transcription activation parameters at ara pBAD. *J. Mol. Biol.* **258**, 14–24 (1996).
21. Jair, K.W., Fawcett, W.P., Fujita, N., Ishihama, A. & Wolf, R.E., Jr. Ambidextrous transcriptional activation by SoxS: requirement for the C-terminal domain of the RNA polymerase alpha subunit in a subset of *Escherichia coli* superoxide-inducible genes. *Mol. Microbiol.* **19**, 307–317 (1996).
22. Li, Z. & Demple, B. Sequence specificity for DNA binding by *Escherichia coli* SoxS and Rob proteins. *Mol. Microbiol.* **20**, 937–945 (1996).
23. Martin, R.G., Gillette, W.K., Rhee, S. & Rosner, J.L. Promoter discrimination by the related transcriptional activators MarA and SoxS: differential regulation by differential binding. *Molec. Microbiol.* **35**, 623–634 (2000).
24. Niland, P., Huhne, R. & Muller-Hill, B. How AraC interacts specifically with its target DNAs. *J. Mol. Biol.* **264**, 667–674 (1996).
25. Reeder, T. & Schleif, R. AraC protein can activate transcription from only one position and when pointed in only one direction. *J. Mol. Biol.* **231**, 205–218 (1993).
26. Brunelle, A. & Schleif, R. Determining residue-base interactions between AraC protein and aral DNA. *J. Mol. Biol.* **209**, 607–622 (1989).
27. Hendrickson, W. & Schleif, R. A dimer of AraC protein contacts three adjacent major groove regions of the aral DNA site. *Proc. Natl Acad. Sci. USA* **82**, 3129–3133 (1985).
28. Martin, R.G., Jair, K.W., Wolf, R.E., Jr. & Rosner, J.L. Autoactivation of the marRAB multiple antibiotic resistance operon by the MarA transcriptional activator in *Escherichia coli*. *J. Bacteriol.* **178**, 2216–2223 (1996).
29. Koh, Y.S., Chung, W.-H., Lee, J.-H. & Roe, J.-H. The reversed SoxS-binding site upstream of the *ribA* promoter in *Escherichia coli*. *Mol. Gen. Genet.* **261**, 374–380 (1999).
30. Nunoshiba, T., Hidalgo, E., Li, Z. & Demple, B. Negative autoregulation by the *Escherichia coli* SoxS protein: a dampening mechanism for the soxRS redox stress response. *J. Bacteriol.* **175**, 7492–7494 (1993).
31. Wood, T.I. *et al.* Interdependence of the position and orientation of SoxS binding sites in the transcriptional activation of the class I subset of *Escherichia coli* superoxide-inducible promoters. *Mol. Microbiol.* **34**, 414–430 (1999).
32. Doublie, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530 (1997).
33. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 461–472 (1997).
34. Bailey, S. The CCP4 suite: programs for protein crystallography. *Acta Crystallog. D* **52**, 30–42 (1994).
35. Abrahams, J.P. & Leslie, A.G.W. Methods used in the structure determination of bovine mitochrondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
36. Brunger, A.T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
37. Jones, T.A. *O: the manual* (, Uppsala, Sweden; 1992) http://kaktus.kemi.aau.dk
38. Brunger, A.T. The free R-value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–474 (1992).
39. Evans, S.V. SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graph.* **11**, 134–138, (1993).
40. Barton, G.J. ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37–40 (1993).