Amino Acid-DNA Contacts by RhaS: an AraC Family Transcription Activator

PRASANNA M. BHENDE AND SUSAN M. EGAN*

Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas 66045

RhaS, an AraC family protein, activates *rhaBAD* transcription by binding to *rhaI*, a site consisting of two 17-bp inverted repeat half-sites. In this work, amino acids in RhaS that make base-specific contacts with *rhaI* were identified. Sequence similarity with AraC suggested that the first contacting motif of RhaS was a helix-turn-helix. Assays of *rhaB-lacZ* activation by alanine mutants within this potential motif indicated that residues 201, 202, 205, and 206 might contact *rhaI*. The second motif was identified based on the hypothesis that a region of especially high amino acid similarity between RhaS and RhaR (another AraC family member) might contact the nearly identical DNA sequences in one major groove of their half-sites. We first made targeted, random mutations and then made alanine substitutions within this region of RhaS. Our analysis identified residues 247, 248, 250, 252, 253, and 254 as potentially important for DNA binding. A genetic loss-of-contact approach was used to identify whether any of the RhaS amino acids in the first or second contacting motif make base-specific DNA contacts. In motif 1, we found that Arg202 and Arg206 both make specific contacts with bp -65 and -67 in $rhaI_1$, and that Arg202 contacts -46 and Arg206 contacts -48 in $rhaI_2$. In motif 2, we found that Asp250 and Asn252 both contact the bp -79 in $rhaI_1$. Alignment with the recently crystallized MarA protein suggest that both RhaS motifs are likely helix-turn-helix DNA-binding motifs.

The AraC family of transcription activators (sometimes also called the AraC/XylS family) was found to consist of more than 100 members in the last published homology search (performed in March 1997) (13). Activators in this family are found in a wide variety of bacterial species. Many members of this family regulate expression of genes involved in carbon metabolism, stress responses, or pathogenesis. Examples of proteins that regulate carbon metabolism (and the carbon source and species they are found in) are AraC (L-arabinose, *Escherichia coli*), RhaS and RhaR (L-rhamnose, *E. coli*), XylS (benzoates, *Pseudomonas putida*), and MelR (melibiose, *E. coli*) (2, 14, 33, 36). Some of the proteins that regulate stress response genes are SoxS (oxidative stress, *E. coli*), MarA (multiple antibiotic resistance, *E. coli*), Ada (alkylating agents, *E. coli*), and Rob (originally identified as binding to *oriC* but cross-activates SoxS- and MarA-regulated genes, *E. coli*) (6, 18–21, 30, 39). Finally, some AraC family proteins that regulate virulence factors in pathogens include CfaD (fimbriae, enterotoxigenic *E. coli*) and VirF (invasion proteins, *Shigella dysenteriae*) (11, 17, 31, 32).

The RhaS protein, one of the AraC family members, activates expression of the *E. coli rhaBAD* and *rhaT* operons (10, 38). *rhaT* encodes the L-rhamnose transport protein (35). *rhaBAD* encodes the enzymes required for catabolism of the sugar L-rhamnose (25, 28); hence RhaS binding and activation of *rhaBAD* occurs only in the presence of L-rhamnose (9, 10). In the absence of L-rhamnose, there is apparently insufficient active RhaS to saturate the *rhaBAD* promoter, resulting in a lag in *rhaBAD* mRNA accumulation after L-rhamnose addition (10). RhaR, another AraC family protein (12, 13, 29, 36), activates *rhaSR* expression (37). Although RhaS and RhaR are

encoded in the same operon, there apparently is sufficient RhaR to saturate the *rhaSR* promoter in the absence of L-rhamnose, as there is no detectable lag in *rhaSR* mRNA accumulation after L-rhamnose induction (10). As a result, L-rhamnose induction of *rhaBAD* expression first requires induction of *rhaSR* expression, which leads to accumulation of RhaS and finally activation of *rhaBAD* expression.

Egan and Schleif (9) used DNase I footprinting and point mutations to locate the DNA-binding site for RhaS (*rhaI*) to a region spanning -32 to -81 relative to the *rhaBAD* transcription start site (Fig. 1). The symmetry of important single-base mutations in *rhaI* led to the prediction that the RhaS binding site is an inverted repeat of two 17-bp half-sites ($rhaI_1$ and $rhaI_2$) separated by 16 bp of uncontacted DNA. Based on the fact that AraC binds to DNA as a dimer (16) and the size of the RhaS binding site, it is predicted that RhaS binds to *rhaI* as a dimer, with each RhaS monomer contacting one half-site. When viewed from one face of the DNA helix, each *rhaI* half-site consists of two major grooves, separated by a minor groove. All of the base pairs found important for RhaS binding lie in the major grooves of *rhaI* (9), suggesting that RhaS makes all of its sequence-specific DNA contacts within the major grooves (Fig. 1). This is in agreement with the previous findings that AraC and RhaR also interact with their DNA binding sites within two major grooves of each half-site (3, 37). We expected that in order to contact two major grooves, each RhaS monomer would possess two DNA-contacting motifs.

One DNA-binding motif in AraC has been predicted to be a helix-turn-helix (H-T-H) motif consisting of residues 196 to 215 (3). Ser208 and His212 of this motif have been shown to make base-specific contacts with DNA (3, 27). Asp256 of AraC has also been shown to be important for DNA sequence recognition and was predicted to be a part of a second H-T-H motif (H-T-H 2) (27). Recently Rhee et al. (30) determined the crystal structure of the first AraC family member, MarA. MarA is one of relatively few AraC family members that con-

* Corresponding author. Mailing address: 8031 Haworth Hall, Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66045. Phone: (785) 864-4294. Fax: (785) 864-5294. E-mail: sme @kuhub.cc.ukans.edu.
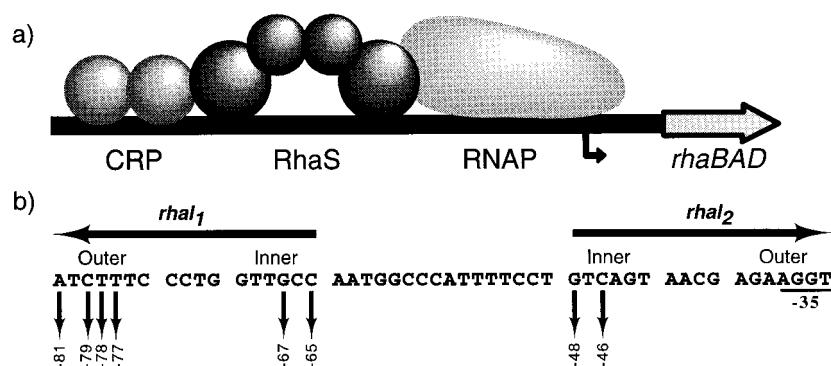
FIG. 1. (a) Model for transcription activation at *rhaBAD*, showing the two activator proteins, cyclic AMP receptor protein (CRP) and RhaS, and RNA polymerase (RNAP) shown. The bent arrow represents the *rhaBAD* transcription start site. (b) DNA sequence of the RhaS binding site (*rhaI*). The two half-sites, *rhaI*₁ and *rhaI*₂, are inverted repeats, as indicated by the arrows above the sequence. Sequences within the half-sites are divided into two major grooves separated by a minor groove, with the outer and inner major grooves indicated. Down arrows indicate bases previously identified as important for interaction with RhaS (9); their positions relative to the *rhaBAD* transcription start site are indicated below. Bases overlapping the −35 region of the promoter were not analyzed in this previous study since their effects on RhaS binding and RNA polymerase binding could not easily be distinguished by the analysis used. Notice that all of the bases important in *rhaI*₁ are identical in *rhaI*₂.

sists of a DNA-binding domain without any dimerization domain, and it activates transcription as a monomer (6, 23). The MarA structure includes two H-T-H DNA-binding motifs; the first (H-T-H 1) aligns with AraC Ser208 and His212, and the second (H-T-H 2) aligns with AraC Asp256.

In this study, we used a genetic analysis to identify RhaS amino acids that make specific contacts with *rhaI*. Our results provide further evidence that AraC family transcription activators utilize two H-T-H motifs to contact DNA. Further, this work is the most extensive functional analysis of the second AraC family H-T-H motif to date. Our data suggest that Arg202 and Arg206 of RhaS, which align with H-T-H 1 of MarA and AraC, make base-specific contacts with *rhaI*. Both of these amino acids contact the inner major grooves of the inverted repeat RhaS binding site (Fig. 1). Our results further show that residues within the 247 to 254 region of RhaS are

crucial for contact with *rhaI*. We found that Asp250 and Asn252, which lie in this region, make specific contacts with the outer major grooves of *rhaI* and are part of H-T-H 2 of RhaS.

MATERIALS AND METHODS

Culture media. Cultures for β-galactosidase assay were grown in 1× MOPS [3-(*N*-morpholino)propanesulfonic acid]-buffered medium (26) (40 mM MOPS, 4 mM Tricine, 0.01 mM FeSO₄, 9.5 mM NH₄Cl, 0.276 mM K₂SO₄, 0.5 μM CaCl₂, 0.528 mM MgCl₂, 50 mM NaCl, 3×10^{-9} M Na₂Mo₄, 4×10^{-7} M H₃BO₃, 3×10^{-8} M CoCl₂, 10^{-8} M CuSO₄, 8×10^{-8} M MnCl₂, 10^{-8} M ZnSO₄, 1.32 mM K₂HPO₄, 10 mM NaHCO₃, 0.2% Casamino Acids, 0.002% thiamine). Ampicillin and tetracycline were used at 125 and 12.5 μg/ml, respectively, when necessary. For other experiments (cloning, strain construction, Ter test [15], etc.), cells were grown in tryptone-yeast extract medium (22) with or without antibiotic or in TB-maltose (0.8% Bacto-Tryptone, 0.5% NaCl, 0.2% maltose).

DNA sequencing. IRD41 dye-labeled primers for DNA sequencing (Table 1) were custom made by LI-COR, Inc. (Lincoln, Neb.). DNA sequences were verified by automated dideoxy sequencing on a LI-COR 4000L sequencer. Se-

TABLE 1. Oligonucleotides used in this study

Designation	Sequence, 5′-3′[a]	Use
898	TGAGTAAAGCTTTTATTGCAGAAAGCCATCCCG	Amplify *rhaS*, downstream
1170	CCGGAATTCTTGTGGTGATGTGATGCTCAC	Amplify *rhaS*, upstream
2064	TCTCTTTCACTGGCTACGCTACATCG	Arg202Ala
2065	CGTACGCTACATGCGCAGCTTAAGCAG	Arg206Ala
2066	CAATTTTCTCTTTCAGCGCGTACGCTACAT	Leu201Ala
2067	CCTATCGCTGTGGATTCAGCGACAGTAACCACTTTTCGACGCTTT[b]	H-T-H 2 random mutations
2068	ATGACCGTATTACATAGTGTGGAT[c]	*rhaS* sequencing
2069	TTATTGCAGAAAGCCATCCCGTCC[c]	*rhaS* sequencing
2073	GGATTCAGCGCCAGTAACCAC	Asp250Ala
2074	TGGTTGCACAGATGGAACAGC[c]	*rhaS* sequencing
2075	GTTGAGACGTGATGCGCTGTT[c]	*rhaS* sequencing
2087	TTCAGCGACAGTGCCCACTTT	Asn252Ala
2088	TGTGGAGCCAGCGACAGTAAC	Phe248Ala
2089	AGTAACGCCTTTTCGACGCTTTTT	His253Ala
2090	AACCACGCTTCGACGCTTTTT	Phe254Ala
2091	AGTAACCACTTTGCGACGCTTT	Ser255Ala
2093	TATCGCGCTGGATTCAGC	Cys246Ala
2120	CGTACGCTAGCTCGGCAGCTTA	His205Ala
2125	CCACTTTTCGGCGCTTTTTCG	Thr256Ala

[a] For oligonucleotides used for PCR amplification, sequences complementary to *rhaS* are underlined; for those used for site-directed mutagenesis, positions of mutations are underlined.

[b] During synthesis, the wild-type nucleotide as well as a low concentration of each of the other three nucleotides was added at each of the underlined positions, resulting in a pool of oligonucleotides with an average of approximately 1.5 mutations each.

[c] IRD41 dye labeled for use in a LI-COR automated sequencer.

TABLE 2. Strains, phage, and plasmids used in this study

Strain, phage, or plasmid	Genotype	Reference or source
E. coli strains		
ECL116	F⁻ Δ*lacU169 endA hsdR thi*	1
SME1082	ECL116 Δ*rhaS*[a]	Laboratory collection
SME1212	SME1082 λSME103 C(−46)G[b]	This study
SME1213	SME1082 λSME103 T(−78)G[b]	This study
SME1214	SME1082 λSME103 C(−65)T[b]	This study
SME1221	SME1082 λSME103 T(−77)G[b]	This study
SME1222	SME1082 λSME103 wt[c]	This study
SME1724	SME1082 λSME103 G(−48)T[b]	This study
SME1725	SME1082 λSME103 A(−81)C[b]	This study
SME1726	SME1082 λSME103 G(−67)A[b]	This study
SME1733	SME1082 λSME103 C(−79)A[b]	This study
Phage		
λRS45	*bla'-lacZ*ₛc *att⁺ int⁺ imm²¹*	34
λSME103	λRS45 Φ(*rhaB-lacZ*)Δ110	10
Plasmids		
pRS414	Apʳ, *'lacZ lacY⁺ lacA⁺*	34
pALTER-1	Apˢ Tetʳ *lacZ* f1 ori	Promega Corp.
pSE101	Apʳ *rhaBAD⁺ rhaSR⁺*	10
pSE105	pRS414 Φ(*rhaB-lacZ*) (wt)	10
pSE159	Apʳ pALTER-1 *rhaS* (wt)	This study
pSE160	Tetʳ pALTER-1 *rhaS* (wt)	This study
pSE165	pSE160 (Leu201Ala)	This study
pSE166	pSE159 (Arg202Ala)	This study
pSE167	pSE159 (His205Ala)	This study
pSE168	pSE160 (Arg206Ala)	This study
pSE169	pSE159 (Cys246Ala)	This study
pSE170	pSE160 (Gly247Ala)	This study
pSE171	pSE159 (Phe248Ala)	This study
pSE172	pSE160 (Asp250Ala)	This study
pSE173	pSE159 (Asn252Ala)	This study
pSE174	pSE159 (His253Ala)	This study
pSE175	pSE159 (Phe254Ala)	This study
pSE176	pSE159 (Ser255Ala)	This study
pSE177	pSE159 (Thr256Ala)	This study
pSE178	pSE159 (Gly247Arg)	This study
pSE179	pSE160 (Phe248Val)	This study
pSE180	pSE160 (Ser249Gly)	This study
pSE181	pSE160 (Ser251Gly)	This study
pSE182	pSE159 (Asn252His)	This study
pSE183	pSE159 (Asn252Ser)	This study
pSE184	pSE160 (Asn252Ile)	This study
pSE185	pSE160 (His253Gln)	This study

[a] Construction of this Δ*rhaS* allele is described in reference 10.
[b] Original mutations described in reference 9.
[c] wt, wild type.

quencing reactions were performed with a Thermo Sequenase fluorescent-labelled primer cycle sequencing kit from Amersham Life Science.

Plasmids, phage, and strains. Wild-type *rhaS* was amplified by PCR using primers 1170 and 898 and plasmid pSE101 as the template. The PCR product was digested at the *Eco*RI site in 1170 and the *Hin*dIII site in 898 and cloned between the *Eco*RI and *Hin*dIII sites of Tetʳ pALTER-1, resulting in pSE160. An Apʳ version of the clone (pSE159) was made by using Promega Altered Sites II in vitro mutagenesis system. The DNA sequence of the cloned *rhaS* gene in both pSE159 and pSE160 was verified by DNA sequencing on both strands.

All strains used were derivatives of ECL116 (1) (Table 2). Translational *rhaB-lacZ* fusions with point mutations within *rhaI* (9) were initially present on derivatives of plasmid pSE105 (Table 2). These fusions were transferred to phage λRS45 (λ*imm²¹*) by in vivo recombination (34) to generate recombinant λ phages (Table 2). The final strains were constructed by transducing SME1082 with the corresponding recombinant λ phage. Lysogens were selected by spreading the transduction mixture on a plate carrying a lawn of λgt30 (λ*imm²¹*). Lysogens were differentiated from λ phage-resistant cells by their sensitivity to a heteroimmune phage upon cross-streaking. Single lysogens were identified by the Ter test (15). In the case of (*rhaB-lacZ*), fusions with single point mutations at −65 and −78, a lysogen was first made by transducing ECL116 with the corresponding recombinant λ phage and confirmed by the Ter test (15). The single lysogen was then introduced into SME1082 by phage P1-mediated generalized transduction (24).

β-Galactosidase assay. Starter cultures of strains to be assayed were grown in tryptone-yeast extract broth containing appropriate antibiotic for approximately 7 h at 37°C; 40 µl of starter culture was then inoculated into 2.5 ml of overnight

TABLE 3. Alanine substitutions in RhaS H-T-H 1 analyzed at wild-type *rhaB-lacZ*

RhaS derivative	% of wild-type activity[a] (sp act [U])
Leu201Ala	48 (238)
Arg202Ala	4 (14)
His205Ala	28 (97)
Arg206Ala	0.1 (0.7)

[a] β-Galactosidase specific activity was measured from a single-copy *rhaB-lacZ* fusion in a *rhaS* deletion strain transformed with plasmids encoding either a wild-type or mutant form of RhaS. Cultures were grown in MOPS growth medium containing glycerol, L-rhamnose, and appropriate antibiotic. The activity of RhaS mutants is represented as the percentage of the activity of wild-type RhaS. Standard errors were less than 10% of the average units.

medium and grown for approximately 17 h. Overnight medium consisted of 1× MOPS medium containing 0.04% glycerol as the carbon source and appropriate antibiotic. A 100-µl volume of the overnight culture was inoculated into 10 ml of growth medium. Growth medium consisted of 1× MOPS medium containing 0.4% glycerol as the carbon source with 0.2% L-rhamnose added as inducer in all cases and appropriate antibiotic. Cultures were grown at 37°C with vigorous shaking in baffled, 125-ml Erhlenmeyer flasks to an A_{600} of approximately 0.4. Growth medium cultures were then centrifuged, and the cells were resuspended in Z buffer (24). β-Galactosidase activity was determined as described by Miller (24) except that incubation with substrate *o*-nitrophenyl-β-D-galactopyranoside was at room temperature. Specific activities were averaged from at least three independent assays, with two replicates in each assay.

Mutagenesis of *rhaS*. Oligonucleotide primers used for random and site-directed mutagenesis (Table 1) were synthesized by Oligos Etc. Inc. A 45-nucleotide oligonucleotide with degenerate sequence (5) over the region to be mutagenized was used to generate random mutations in the *rhaS* region encoding amino acids 246 to 255 (Table 1). The oligonucleotide was designed to introduce an average of 1.5 mutations within the 30-bp region and was used in a site-directed mutagenesis procedure (Promega Altered Sites II in vitro mutagenesis system) to create random mutations.

Alanine substitutions within the first and second possible DNA-binding motifs of RhaS were created by site-directed mutagenesis (Promega Altered Sites II and GeneEditor in vitro mutagenesis systems) (Table 1). Double-stranded DNA template was used to construct all amino acid substitutions except Arg202Ala and Thr256Ala, for which a single-stranded DNA template was used. Both random and alanine substitutions were entirely sequenced on both strands to confirm each mutation and to ensure that there were no additional amino acid changes. One random mutant (Phe248Val) had two silent mutations in addition to the substitution at residue 248, but all other mutants had DNA sequences identical to that of the wild-type *rhaS* clone.

RESULTS

First RhaS DNA-contacting motif. AraC amino acids Ser208 and His212, which correspond to amino acids 2 and 6 of the DNA recognition helix of an H-T-H motif, have been shown to make specific contacts with base pairs within the AraC binding site (*araI*) (3, 27). These AraC amino acids align with RhaS Arg202 and Arg206. Hence, we created alanine substitutions at Arg202 and Arg206 of RhaS. Amino acid 1 of the recognition helix often is involved in making specific DNA contacts, and so we also made an alanine substitution at Leu201. Further, it has been recently shown that Gln45 of MarA makes specific contact with DNA (30). Since MarA Gln45 aligns with RhaS His205, we also created an alanine substitution at His205 of RhaS.

We measured the ability of each of the above RhaS mutants to activate transcription from a wild-type Φ(*rhaB-lacZ*) fusion to determine whether the mutant residues might be important for RhaS to contact DNA (Table 3). All four of the mutations tested (Leu201Ala, Arg202Ala, His205Ala, and Arg206Ala) were at least twofold defective in transcription activation at the wild-type Φ(*rhaB-lacZ*) fusion (Table 3). Arg202Ala and Arg206Ala had especially large defects. These data are consistent with the hypothesis that Leu201, Arg202, His205, and Arg206 are part of a RhaS H-T-H DNA-binding motif in-
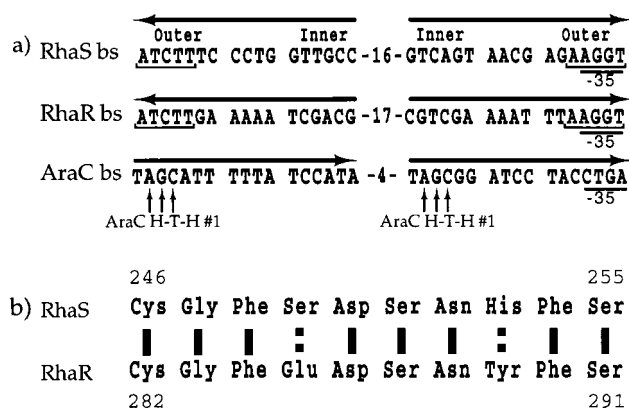
a) RhaS bs

b) RhaS

FIG. 2. (a) Comparison of the DNA-binding sites (bs) recognized by RhaS, RhaR, and AraC. Horizontal arrows indicate the half-sites and their relative orientations. Notice that the downstream half-sites are all in the same relative orientation whereas the upstream half-sites are not. Bracketed bases are conserved in the RhaS and RhaR DNA-binding sites. The outer and inner major grooves of the RhaS binding site are indicated. The −35 regions of the promoters are shown. Vertical arrows indicate amino acid-base pair contacts between AraC H-T-H 1 and its DNA-binding site. (b) Sequence of amino acids 246 to 255 of RhaS, showing high similarity with the aligned region of RhaR (amino acids 282 to 291). Solid lines indicate identical amino acids; broken lines represent similar amino acids.

volved in contacting *rhaI*, although other explanations are possible on the basis of this result alone.

Random mutagenesis of the second possible RhaS DNA-contacting motif. As described above, RhaS is expected to have two DNA-binding motifs. It seemed likely that the RhaS H-T-H involving amino acids 201 to 206 would contact the inner major grooves of *rhaI* (Fig. 2a), based on the expectation that the promoter proximal RhaS and AraC monomers would bind to DNA in the same relative orientation (and possibly make conserved contacts with RNA polymerase). Egan and Schleif (9) previously observed that the outer major grooves of *rhaI* are nearly identical to the corresponding major grooves of the site for RhaR binding (Fig. 2a). This suggests that RhaS and RhaR may share a conserved DNA-binding motif. We found a 10-residue region which was highly conserved in RhaS (residues 246 to 255) and RhaR (residues 282 to 291) (Fig. 2b). Eight of the ten amino acids in this region are identical whereas the other two are similar—a degree of sequence similarity that is unusually high, even for the very similar RhaS and RhaR proteins. Therefore, we hypothesized that residues 246 to 255 might be part of the second DNA-binding motif of RhaS.

We used an oligonucleotide with degenerate sequence (2067 [Table 1]) in a site-directed mutagenesis procedure to create random mutations in this conserved region of RhaS. All of the mutations generated were analyzed, without any functional selection. We isolated mutations at 7 of the 10 amino acid residues in the region (Table 4). As we had already constructed an alanine substitution at amino acid 250 of RhaS (see below), we did not assay the random substitution that we isolated at position 250 (Asp250Glu). To test if RhaS residues from positions 246 to 255 might contact *rhaI*, we first analyzed the abilities of the random mutants to activate transcription from a wild-type Φ(*rhaB-lacZ*) fusion (Table 4). With the exception of Ser249Gly and Ser251Gly, all of the random mutants were considerably defective in activating transcription from wild-type Φ(*rhaB-lacZ*). Defects greater than 10-fold were found with mutations at amino acids Gly247, Phe248, and Asn252. This suggests that these residues and possibly His253, which

had a threefold defect, might be involved in DNA binding by RhaS. These residues are candidates for the second DNA-binding motif of RhaS.

Alanine substitutions in the second RhaS H-T-H motif. We made individual alanine substitutions at most of the amino acid positions in the 246–255 region. Alanine substitutions were not constructed at Ser249 or Ser251 since random Gly substitutions at these positions had very little effect on RhaS function. We did not construct an alanine substitution at Gly247 since an alanine substitution at this position was isolated in the random mutagenesis described above. While we were in the process of creating and analyzing the alanine substitutions, publication of the MarA crystal structure showed that Arg96 makes specific contacts with DNA (30). Since, MarA Arg96 aligns with RhaS Thr256 (13), we also made an alanine substitution at this position.

We tested these RhaS alanine mutants for the ability to activate transcription at wild-type Φ(*rhaB-lacZ*). Alanine mutants at Gly247, Phe248, Asp250, Asn252, His253, and Phe254 were all significantly defective for transcription activation from the wild-type fusion (Table 5). This result supports the hypothesis that these amino acids might be part of a second DNA-binding motif in RhaS, which we predict to be an H-T-H motif based on alignment with MarA (30).

Base-specific contacts by RhaS. To test whether base-specific contacts are made by any of the amino acids thought to be

TABLE 4. Random mutations in RhaS H-T-H 2 analyzed at wild-type *rhaB-lacZ*

RhaS derivative	% of wild-type activity[a] (sp act [U])
Gly247Arg	0.9 (3.8)
Phe248Val	3.3 (17)
Ser249Gly	93 (536)
Ser251Gly	87 (505)
Asn252His	0.4 (2.0)
Asn252Ser	37 (184)
Asn252Ile	0.3 (2.0)
His253Gln	32 (185)

[a] β-Galactosidase specific activity was measured from a single-copy *rhaB-lacZ* fusion in a *rhaS* deletion strain transformed with plasmids encoding either a wild-type or mutant form of RhaS. Cultures were grown in MOPS growth medium containing glycerol, L-rhamnose, and appropriate antibiotic. The activity of RhaS mutants is represented as the percentage of the activity of wild-type RhaS. Standard errors were less than 21% of the average units.

TABLE 5. Alanine substitutions in RhaS H-T-H 2 analyzed at wild-type *rhaB-lacZ*

RhaS derivative	% of wild-type activity[a] (sp act [U])
Cys246Ala	66 (174)
Gly247Ala	6.2 (33)
Phe248Ala	0.9 (2.3)
Asp250Ala	8.3 (44)
Asn252Ala	31 (80)
His253Ala	48 (153)
Phe254Ala	1.2 (3.2)
Ser255Ala	80 (210)
Thr256Ala	61 (208)

[a] β-Galactosidase specific activity was measured from a single-copy *rhaB-lacZ* fusion in a *rhaS* deletion strain transformed with plasmids encoding either a wild-type or mutant form of RhaS. Cultures were grown in MOPS growth medium containing glycerol, L-rhamnose, and appropriate antibiotic. The activity of RhaS mutants is represented as the percentage of the activity of wild-type RhaS. Standard errors were less than 8% of the average units.

important in DNA binding, we used a genetic loss-of-contact approach (7, 8). The rationale behind the genetic loss-of-contact approach is that a protein with a single amino acid substitution will lose the ability to discriminate between the wild-type and mutant base pairs only at the position or positions in the DNA that are normally contacted by the substituted amino acid. As a result, any DNA positions normally contacted by the substituted amino acid can be mutated with relatively little effect, while mutations at other DNA positions will have large effects. This analysis will determine whether some of the amino acids that we have identified are directly involved in DNA binding and also identify the base pair positions contacted by such base-specific amino acids.

Each of the alanine substitutions found to be at least twofold defective at wild-type Φ(*rhaB-lacZ*) was tested at eight derivatives of Φ(*rhaB-lacZ*) (Fig. 3). The derivatives contained single point mutations at the important base positions in *rhaI*₁ and in the upstream major groove of *rhaI*₂ (Fig. 1). Within RhaS H-T-H 1, we found that Leu201Ala and His205Ala gave considerably lower activity at all of the mutant Φ(*rhaB-lacZ*) fusions compared with wild-type Φ(*rhaB-lacZ*) (Fig. 3). This indicates that Leu201 and His205 do not make any specific contacts with *rhaI*. In contrast, point mutations at −65 and −67 within *rhaI*₁ had relatively little effect on activation by both Arg202Ala and Arg206Ala (Fig. 3). In *rhaI*₂, the point mutation at −46 had relatively little effect on activation by Arg202Ala and the point mutation at −48 had relatively little effect on activation by Arg206Ala. These data suggest that Arg202 contacts the base pairs at −46, −65, and −67, and Arg206 contacts the base pairs at −48, −65, and −67. This result strongly suggests that Arg202 and Arg206 are both directly involved in DNA binding and do not simply have indirect effects on DNA binding by RhaS.

We also tested the defective mutants in the second H-T-H motif of RhaS at mutant Φ(*rhaB-lacZ*) fusions to determine if some or all of them make specific contacts with *rhaI* (Fig. 3). Gly247Ala, Phe248Ala, His253Ala, and Phe254Ala all gave significantly lower β-galactosidase specific activities from all mutant Φ(*rhaB-lacZ*) fusions compared with the wild-type fusion. This result suggests that these amino acids do not make any specific contacts with *rhaI*. With Asp250Ala and Asn252Ala, on the other hand, a point mutation at −79 had relatively little effect on activation. We conclude that both Asp250 and Asn252 are directly involved in DNA binding and make specific contacts with the base pair at position −79 within *rhaI*. In addition, Asn252Ala gave a significantly higher level of activity with the point mutation at −77 than the remaining base changes in *rhaI*, suggesting that Asn252 may make a second contact with the base pair at position −77.

DISCUSSION

In this work, we set out to determine the amino acids in the RhaS protein that contact *rhaI*, as well as the specific bases in *rhaI* that are contacted by those amino acids. Each of the *rhaI* half-sites is 17 bp long (Fig. 1), and therefore the face of the DNA helix likely to be contacted by RhaS will consist of two major grooves of DNA (9). This suggests that the RhaS protein contains two distinct DNA-contacting motifs in each monomer.

H-T-H 1 of RhaS is a DNA-binding motif. We constructed alanine substitutions at four amino acid positions within the recognition helix of a RhaS H-T-H DNA-binding motif proposed based on alignment with AraC (3). We found that mutations in the first (Leu201), second (Arg202), fifth (His205), and sixth (Arg206) amino acids of the recognition helix re-

sulted in defects in RhaS activation from the wild-type *rhaBAD* promoter (Table 3). This result is consistent with the hypothesis that these amino acids may be involved in sequence-specific DNA contacts by RhaS. As discussed below, our results indicate that this motif contacts the inner major grooves of the *rhaI* inverted repeat.

H-T-H 2 of RhaS is also a DNA-binding motif. At the time we began these experiments, the second DNA-contacting motif had not been identified in any AraC family proteins. We hypothesized that the RhaS and RhaR proteins might use a conserved DNA-contacting motif to recognize the conserved outer major grooves of their binding sites (Fig. 2). We identified amino acids 246 to 255 of RhaS as a region having unusually high sequence similarity with RhaR and proposed that this region might be a part of the second DNA-contacting motif. We found that random substitutions at positions Gly247, Phe248, Asn252, and His253 resulted in a defect in the ability of RhaS to activate expression from the wild-type *rhaBAD* promoter (Table 4). Ala mutants at all of the amino acids from residues 247 to 254, except for 249 and 251 (which were tested as random substitutions only), were found to be at least twofold defective for activation from the *rhaBAD* promoter, and most were more than 10-fold defective (Table 5). We propose that this region is part of a second DNA-binding motif of RhaS that, based on alignment with the crystal structure of MarA (30), is likely an H-T-H motif.

Identification of DNA contacts by RhaS. Further analysis of the mutants that we constructed in the two H-T-H motifs was performed by using a genetic loss-of-contact approach (outlined in Results) (7, 8). A genetic approach was chosen for initial characterization of these mutants since it is more informative than comparison of binding affinities of the mutants to wild-type and much easier than in vitro missing-contact probing (3, 4). In addition, a genetic approach was particularly useful in this case since the RhaS protein is severely insoluble (9).

Using the genetic loss-of-contact approach, we have determined that 4 of the 10 amino acids identified as potentially important for DNA binding by RhaS make specific contacts with base pairs in *rhaI*. In H-T-H 1, amino acids Arg202 and Arg206 were each found to contact the base pairs at positions −65 and −67 in *rhaI*₁. The contacts in *rhaI*₂ were not exactly symmetric, with Arg202 contacting only −46 and Arg206 contacting only −48 (Fig. 1 and 4). Since the contacted bases are identical in both half-sites, the asymmetry in the contacts is likely to be due to either the DNA context of the contacted bases or the influence of proteins that may contact RhaS (cyclic AMP receptor protein and RNA polymerase). Thus, RhaS H-T-H 1 contacts the inner major grooves of the *rhaI* inverted repeat site (Fig. 4), consistent with the expectation that the promoter proximal monomers of RhaS and AraC bind in the same relative orientation.

In RhaS H-T-H 2, Asp250 and Asn252 are both predicted to make base-specific contact with the base pair at position −79 of *rhaI*₁. In addition, our results somewhat weakly suggest that Asn252 may also interact with the base pair at position −77. These base pairs are within the outer major groove of the inverted repeat *rhaI* site, consistent with the finding that H-T-H 1 contacts the inner major grooves. The important bases for RhaS binding in the outer major groove of *rhaI*₂ have not been identified due to their overlap with the −35 region of the promoter. Since the important bases in *rhaI*₁ are identical in *rhaI*₂, we initially predicted that the contacts would be symmetric; however, the finding of asymmetry in the inner major grooves suggests that further investigation is necessary to identify the contacts in this region.
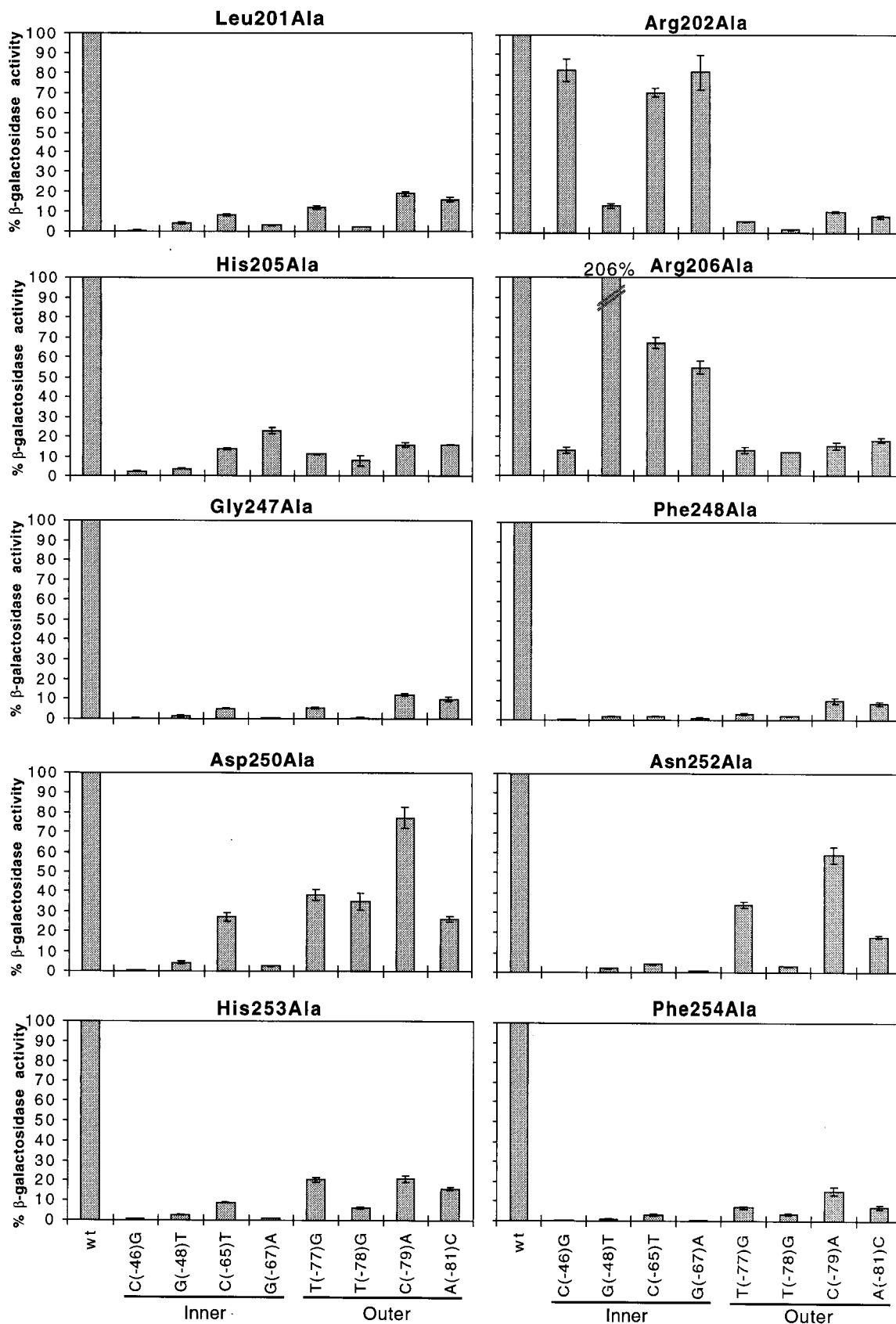
FIG. 3. Alanine substitutions in RhaS H-T-H 1 and 2 analyzed at mutant *rhaB-lacZ* fusions. The *x* axis (labeled at the bottom) represents either the wild-type *rhaBAD* promoter or the position of point mutations in *rhaI* found to be important for RhaS binding (Fig. 1). Locations of the point mutations in either the inner or outer major grooves of *rhaI* are indicated. The *y* axis (labeled on the left) represents the percent β-galactosidase specific activity for each RhaS alanine mutant at mutant *rhaB-lacZ* compared with the same mutant protein at the wild-type *rhaB-lacZ* fusion. The first bar in each graph represents the RhaS alanine mutant protein assayed at the wild-type *rhaB-lacZ* promoter and is set to 100%. Error bars are shown.

```
                 Asp250              Arg206          Arg206
RhaS at          ┌──┐               ┌──┐            ┌──┐
rhaBAD:      -81 ATCTTTC CCTG GTTGCC-16bp-GTCAGT AACG AGAAGGT -32
                 TAGAAAG GGAC CAACGG      CAGTCA TTGC TCTTCCA
                 └──┘              └──┘          └──┘
                 Asn252            Arg202        Arg202

                 Thr95 Arg96        Gln45 Arg46
MarA at          ┌──┐ ┌──┐         ┌──┐ ┌──┐
marRAB:      -70 TTTAGCA AAAC GTGGCA -54
                 AAATCGT TTTG CACCGT
                 └──┘ └──┘        └──┘
                 Gln91 Thr93        Trp42
                 Gln92
                      H-T-H#2      H-T-H#1    H-T-H#1
```
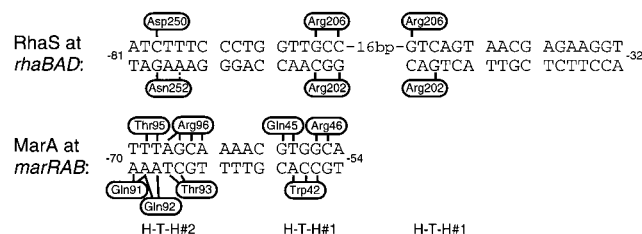
FIG. 4. Summary of specific amino acid-base pair contacts by RhaS and comparison with MarA DNA contacts. DNA half-sites are divided into two major grooves and an intervening minor groove. Vertical and angled black lines represent amino acid-base pair contacts. The broken bar indicates a less conclusive DNA contact by RhaS. The side of the DNA sequence on which each amino acid is represented is arbitrary and is not meant to represent the base within the pair that is contacted. For MarA, the base(s) within each pair that is contacted is known (30); however, this information is not represented for simplicity and because the corresponding information is not known for RhaS. Relative positions of H-T-H 1 and 2 are shown below the two DNA sequences.

Comparison with MarA structure. Both of the amino acids in H-T-H 1 of RhaS that make sequence-specific DNA contacts align with MarA residues also predicted to make specific DNA contacts. Arg202 of RhaS aligns with MarA Trp42. One of the base pairs contacted by each of these amino acids is a G-C base pair in the analogous position in the two binding sites (−56 at *marRAB* and −67 at *rhaBAD*), while each of these residues also makes additional, different contacts (Fig. 4). Arg206 of RhaS aligns with Arg46 of MarA. In this case, the identical amino acid contacts the identical base pair in the analogous position in the two binding sites (G-C at −56 in *marRAB* and −67 in *rhaBAD*), and each amino acid also contacts a C-G base pair (−55 at *marRAB* and −65 at *rhaBAD*). MarA has a third amino acid in H-T-H 1 that contacts DNA (Gln45). MarA Gln45 aligns with RhaS His205, which was found to be important for DNA binding by RhaS but not specific for any base positions. It seems likely that only Arg202 and Arg206 of RhaS H-T-H 1 make DNA contacts since two amino acid partners have been identified for each of the two important base pairs in the inner major grooves.

Asn252 of RhaS aligns with Gln92 of MarA, which also makes specific DNA contacts (Fig. 4). Each of these amino acids contacts one base pair at analogous positions in the two binding sites (−68 at *marRAB* and −79 at *rhaBAD*), as well as a possible second contact by Asn252 (at −77) and one additional contact by Gln92 (at −69). Asp250 of RhaS is the only DNA-contacting amino acid that does not align with a contacting amino acid in MarA. In fact, the alignment of RhaS and MarA (13) predicts that Asp250 lies within the turn prior to the recognition helix of H-T-H 2. There are no DNA contacts (neither specific nor phosphate backbone) by amino acids in this turn of MarA. However, Asp250 of RhaS aligns with Asp256 of AraC, which has also been found important in contacting DNA (27). Interestingly, more than one-fourth of the proteins aligned by Gallegos et al. (13) have an aspartic acid in this position, suggesting that other family members may use this amino acid to contact DNA as well.

There are four amino acids in MarA H-T-H 2 which make base-specific DNA contacts and align with RhaS amino acids that do not make contacts (Fig. 4). These amino acids align with RhaS amino acids judged not to be important for DNA binding (Ser251, Ser255, and Thr256) or found to be important for DNA binding but not base specific (His253). It is not surprising that MarA, which binds to DNA as a monomer (23), has more amino acid-DNA contacts per monomer than RhaS, which is predicted to bind as a dimer. Our current analysis

shows an equal number of amino acids in the MarA monomer and the RhaS dimer that make DNA contacts; however, the total number of base-specific contacts made by the MarA monomer (20 contacts) is greater than the number (12 to 14) made by the RhaS dimer. There is reason to believe that the binding strength of these two proteins may be more similar than the number of contacts suggests. First, we do not think that we have identified all of the DNA contacts made by RhaS with the outer major groove of *rhaI* (see below). Second, MarA would be expected to need more than twice the binding strength of each RhaS monomer to make up for the cooperativity gained by RhaS binding as a dimer.

It is not surprising to find that (i) three of the five base pairs for which we conclusively identified contacts were contacted by two different amino acids and (ii) two of the amino acids in RhaS contact two base pairs separated by an uncontacted base pair. The DNA contacts made by MarA are in accord with both of these findings. There are five different examples in MarA of base pairs contacted by two amino acids, and even one example of a base pair contacted by three amino acids (30) (Fig. 4). Further, two of the amino acids in MarA (Trp42 and Arg96) contact three and four adjacent base pairs, respectively, although with Arg96 there is a water molecule involved in one of the contacts (30).

Have all of the base-specific contacts been identified? Curiously, at least two (or three if −77 is included) important base pairs in the outer major grooves of *rhaI* do not yet have identified amino acid partners. We have tested all of the amino acids ranging from the last amino acid of the first helix, through the turn, up to the sixth amino acid of the recognition helix, without finding any contacts with these base pairs. It is possible that additional amino acids of RhaS make DNA contacts; however, these amino acids are not located in the most common positions for H-T-H contacts. Alternatively, it is possible that amino acids that we tested make DNA contacts that were not detected by our analysis for some reason.

Nonspecific amino acid contacts. Our analysis identified six amino acids in RhaS that were defective for transcription activation at the *rhaBAD* promoter but that did not show evidence of base-specific DNA contacts. It is possible, although certainly not probable, that some of these amino acids are involved in transcription activation rather than DNA binding by RhaS. It is also possible that some of these amino acids make contacts with the phosphate backbone of the DNA, and therefore are directly involved in DNA binding, but do not make base-specific contacts. Eleven different amino acids in MarA make only phosphate backbone contacts with the DNA (30). Finally, it is possible that some of these amino acid substitutions alter the conformation of RhaS and therefore indirectly affect DNA binding. Inspection on the MarA structure of the amino acids that aligns with each of these nonspecific RhaS amino acids (by using RasMol version 2.6) indicates that all of them except Phe254 are significantly surface exposed and near the DNA backbone. Substitutions at Phe254 most likely alter the conformation of RhaS and thereby indirectly reduce DNA binding. Each of the other nonspecific amino acids could possibly make phosphate backbone contacts.

REFERENCES

1. **Backman, K., Y.-M. Chen, and B. Magasanik.** 1981. Physical and genetic characterization of the *gln A-glnG* region of the *Escherichia coli* chromosome. Proc. Natl. Acad. Sci. USA **78:**3743–3747.
2. **Bourgerie, S. J., C. M. Michan, M. S. Thomas, S. J. W. Busby, and E. I. Hyde.** 1997. DNA binding and DNA bending by the MelR transcription activator protein from *Escherichia coli.* Nucleic Acids Res. **25:**1685–1693.
3. **Brunelle, A., and R. Schleif.** 1989. Determining residue-base interactions between AraC protein and *araI* DNA. J. Mol. Biol. **209:**607–622.
4. **Brunelle, A., and R. Schleif.** 1987. Missing contact probing of DNA-protein interactions. Proc. Natl. Acad. Sci. USA **84:**6673–6676.
5. **Chiang, L. W., I. Kovari, and M. M. Howe.** 1993. Mutagenic oligonucleotide-directed PCR amplification (Mod-PCR): an efficient method for generating random base substitution mutations in a DNA sequence element. PCR Methods Appl. **2:**210–217.
6. **Cohen, S. P., H. Hachler, and S. B. Levy.** 1993. Genetic and functional analysis of the multiple antibiotic resistance (*mar*) locus in *Escherichia coli.* J. Bacteriol. **175:**1484–1492.
7. **Ebright, R. H.** 1991. Identification of amino acid-base pair contacts by genetic methods. Methods Enzymol. **208:**620–640.
8. **Ebright, R. H.** 1985. Use of "loss-of-contact" substitutions to identify residues involved in an amino acid-base pair contact: effect of substitution of Gln18 of Lac repressor by Gly, Ser, and Leu. J. Biomol. Struct. Dyn. **3:**281–297.
9. **Egan, S. M., and R. F. Schleif.** 1994. DNA-dependent renaturation of an insoluble DNA binding protein. Identification of the RhaS binding site at *rhaBAD.* J. Mol. Biol. **243:**821–829.
10. **Egan, S. M., and R. F. Schleif.** 1993. A regulatory cascade in the induction of *rhaBAD.* J. Mol. Biol. **234:**87–98.
11. **Falconi, M., B. Colonna, G. Prosseda, G. Micheli, and C. O. Gualerzi.** 1998. Thermoregulation of *Shigella* and *Escherichia coli* EIEC pathogenicity. A temperature-dependent structural transition of DNA modulates accessibility of *virF* promoter to transcriptional repressor H-NS. EMBO J. **17:**7033–7043.
12. **Gallegos, M.-T., C. Michán, and J. L. Ramos.** 1993. The XylS/AraC family of regulators. Nucleic Acids Res. **21:**807–810.
13. **Gallegos, M.-T., R. Schleif, A. Bairoch, K. Hofmann, and J. L. Ramos.** 1997. AraC/XylS family of transcriptional regulators. Microbiol. Mol. Biol. Rev. **61:**393–410.
14. **González-Pérez, M. M., J. L. Ramos, M. T. Gallegos, and S. Marqués.** 1999. Critical nucleotides in the upstream region of the XylS-dependent TOL meta-cleavage pathway operon promoter as deduced from analysis of mutants. J. Biol. Chem. **274:**2286–2290.
15. **Gottesman, M. E., and M. B. Yarmolinsky.** 1968. The integration and excision of the bacteriophage lambda genome. Cold Spring Harbor Symp. Quant. Biol. **33:**735–747.
16. **Hendrickson, W., and R. Schleif.** 1985. A dimer of AraC protein contacts three adjacent major groove regions of the *araI* DNA site. Proc. Natl. Acad. Sci. USA **82:**3129–3133.
17. **Jordi, B. J. A. M., B. A. M. van der Zeijst, and W. Gaastra.** 1994. Regions of the CFA/I promoter involved in the activation by the transcriptional activator CfaD and repression by the histone-like protein H-NS. Biochimie **76:**1052–1054.
18. **Landini, P., and S. J. Busby.** 1999. The *Escherichia coli* Ada protein can interact with two distinct determinants in the σ^{70} subunit of RNA polymerase according to promoter architecture: identification of the target of Ada activation at the *alkA* promoter. J. Bacteriol. **181:**1524–1529.
19. **Lemotte, P. K., and G. C. Walker.** 1985. Induction and autoregulation of *ada*, a positively acting element regulating the response of *Escherichia coli* K-12 to methylating agents. J. Bacteriol. **161:**888–895.
20. **Li, Z., and B. Demple.** 1996. Sequence specificity for DNA binding by *Escherichia coli* SoxS and Rob proteins. Mol. Microbiol. **20:**937–945.
21. **Li, Z., and B. Demple.** 1994. SoxS, an activator of superoxide stress genes in *Escherichia coli*: purification and interaction with DNA. J. Biol. Chem. **269:**18371–18377.
22. **Maloy, S. R., V. J. Stewart, and R. K. Taylor.** 1996. Genetic analysis of pathogenic bacteria. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
23. **Martin, R. G., K.-W. Jair, R. E. Wolf, Jr., and J. L. Rosner.** 1996. Autoactivation of the *marRAB* multiple antibiotic resistance operon by the MarA transcriptional activator in *Escherichia coli.* J. Bacteriol. **178:**2216–2223.
24. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
25. **Moralejo, P., S. M. Egan, E. Hidalgo, and J. Aguilar.** 1993. Sequencing and characterization of a gene cluster encoding the enzymes for L-rhamnose metabolism in *Escherichia coli.* J. Bacteriol. **175:**5585–5594.
26. **Neidhardt, F. C., P. L. Bloch, and D. F. Smith.** 1974. Culture medium for enterobacteria. J. Bacteriol. **119:**736–747.
27. **Niland, P., R. Huhne, and B. Muller-Hill.** 1996. How AraC interacts specifically with its target DNAs. J. Mol. Biol. **254:**667–674.
28. **Power, J.** 1967. The L-rhamnose genetic system in *Escherichia coli* K-12. Genetics **55:**557–568.
29. **Ramos, J. L., F. Rojo, L. Zhou, K. N. Timmis.** 1990. A family of positive regulators related to the *Pseudomonas putida* TOL plasmid XylS and the *Escherichia coli* AraC activators. Nucleic Acids Res. **18:**2149–2152.
30. **Rhee, S., R. G. Martin, J. L. Rosner, and D. R. Davies.** 1998. A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. Proc. Natl. Acad. Sci. USA **95:**10413–10418.
31. **Sakai, T., C. Sasakawa, and M. Yoshikawa.** 1988. Expression of four virulence antigens of *Shigella flexneri* is positively regulated at the transcriptional level by the 30 kiloDalton *virF* protein. Mol. Microbiol. **2:**589–597.
32. **Savelkoul, P. H. M., G. A. Willshaw, M. M. McConnell, H. R. Smith, A. M. Hamers, B. A. M. van der Zeijst, and W. Gaastra.** 1990. Expression of CFA/I fimbriae is positively regulated. Microb. Pathog. **8:**91–99.
33. **Schleif, R.** 1996. Two positively regulated systems, *ara* and *mal*, p. 1300–1309. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology. ASM Press, Washington, D.C.
34. **Simons, R. W., F. Houman, and N. Kleckner.** 1987. Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. Gene **53:**85–96.
35. **Tate, C. G., J. A. R. Muiry, and P. J. F. Henderson.** 1992. Mapping, cloning, expression, and sequencing of the *rhaT* gene which encodes a novel L-rhamnose-H$^+$ transport protein in *Salmonella typhimurium* and *Escherichia coli.* J. Biol. Chem. **287:**6923–6932.
36. **Tobin, J. F., and R. F. Schleif.** 1987. Positive regulation of the *Escherichia coli* L-rhamnose operon is mediated by the products of tandemly repeated regulatory genes. J. Mol. Biol. **196:**789–799.
37. **Tobin, J. F., and R. F. Schleif.** 1990. Transcription from the *rha* operon p$_{sr}$ promoter. J. Mol. Biol. **211:**1–4.
38. **Via, P., J. Badia, L. Baldoma, N. Obradors, and J. Aguilar.** 1996. Transcriptional regulation of the *Escherichia coli rhaT* gene. Microbiology **142:**1833–1840.
39. **Wu, J., and B. Weiss.** 1991. Two divergently transcribed genes, *soxR* and *soxS*, control a superoxide response regulon of *Escherichia coli.* J. Bacteriol. **173:**2864–2871.