J. Mol. Biol. (1998) **278**, 539–548

JMB

Arm-Domain Interactions in AraC

Beatrice Saviola, Robert Seabold and Robert F. Schleif*

Department of Biology
Johns Hopkins University
3400 N. Charles Street
Baltimore, MD 21218, USA

N-terminal deletions extending beyond the sixth amino acid of the *Escherichia coli* regulator of the L-arabinose operon, AraC, were found to generate constitutive regulatory behavior of the promoter p_{BAD}. Mutagenesis of the DNA coding for the first 20 amino acids of the protein and screening for constitutives yielded mutants across the region whereas screening for mutants that cannot induce p_{BAD}, even in the presence of arabinose, yielded none. These results indicate that the N-terminal arm is not essential for transcription activation, but that it plays an important and active role in holding the system in a non-activating state. Despite the fact that arabinose binds to the N-terminal domain of AraC, mutations were found in the C-terminal domain that weaken the binding of arabinose to the protein. The effects of the mutations could be suppressed by specific mutation in the N-terminal arm or by deletion of the arm. These results, in conjunction with the crystal structures of the N-terminal domain determined in the presence and absence of arabinose, indicate that in the absence of arabinose, the N-terminal arms of the protein bind to the C-terminal DNA binding domains to hold them in a state where the protein prefers to loop. When arabinose is added, the arms are pulled off the C-terminal domains, thereby releasing them to bind to adjacently located DNA half-sites and activate transcription.

**Corresponding author*

Keywords: arm; domain-domain; mutation; activation; affinity

Introduction

In the absence of arabinose, the dimeric regulator of the L-arabinose operon in *Escherichia coli*, AraC, prefers to bind to two half-sites that are separated from one another by several hundred base-pairs, thereby forming a DNA loop, (Wilcox & Meuris, 1976; Dunn *et al*., 1984; Hahn *et al*., 1984; Martin *et al*., 1986; Huo *et al*., 1988; Lee & Schleif, 1989; Seabold & Schleif, 1998). This looping, between half-sites I_1 and O_2 (Figure 1), prevents AraC from occupying the I_2 half-site and helps prevent induction of the *araBAD* promoter, p_{BAD}. Further, it is likely, but not proven, that looping in the absence of arabinose also represses the adjacent promoter p_C, whereas it is largely the direct binding of AraC to the O_1 pair of half-sites that represses p_C in the presence of arabinose (Martin *et al*., 1986; Huo *et al*., 1988; Lobell & Schleif, 1990; X. Zhang *et al*., 1996). When arabinose is added, the preference of AraC to engage in looping interactions is reduced, and instead, the protein prefers to bind to two half-sites that are adjacent to one another along the DNA (Lobell & Schleif, 1990; Carra & Schleif, 1993; Seabold & Schleif, 1998). At p_{BAD} this *cis* binding leads to occupancy of the I_1 and I_2 half-sites that, in conjunction with cyclic AMP receptor protein (Lee *et al*., 1974), stimulates binding of RNA polymerase to the promoter and accelerates open complex formation (Greenblatt & Schleif, 1971; Hendrickson & Schleif, 1984, 1985; Lee *et al*., 1987; Lobell & Schleif, 1990; Reeder & Schleif, 1993; X. Zhang *et al*., 1996; Zhang & Schleif, 1998).

The mechanism by which the binding of arabinose causes AraC to shift from preferring to loop to preferring to bind *cis* should be understandable in terms of the protein's structure. AraC protein consists of two loosely connected domains. The N-terminal domain that both binds arabinose and dimerizes the protein is connected by a flexible linker to the DNA binding domain (Bustos & Schleif, 1993; Eustance *et al*., 1994). The dimerization domain possesses at least some of the determinants necessary for generating the arabinose response because a chimeric AraC-LexA protein, in which the DNA binding domain of LexA replaces the DNA binding domain of AraC, displays an arabinose response, albeit much reduced compared to that of wild-type AraC. Conversely, when the C-terminal DNA binding domain of AraC is fused to the leucine zipper dimerization region from C/EBP, the hybrid protein can bind to DNA sites

0022–2836/98/180539–10 $25.00/0/mb981712

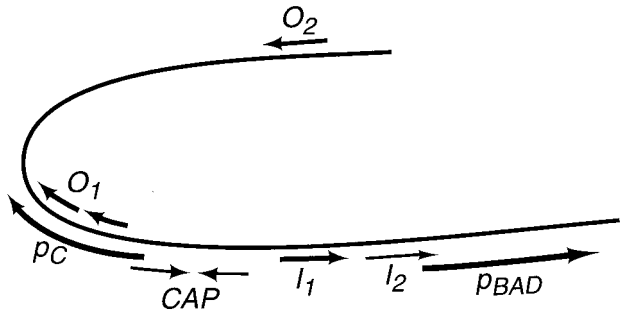

Figure 1. The regulatory region region of the *araCBAD* genes showing the I_1, I_2, and O_2 half-sites and the O_1 pair of half-sites. Occupancy of these sites by AraC, in conjunction with cyclic AMP receptor protein that binds at the CAP site, regulates the activities of both the p_{BAD} and p_C promoters.

specific for AraC and activate transcription from p_{BAD} (Bustos & Schleif, 1993). Thus, the determinants for DNA binding and transcription activation lie within the C-terminal domain.

The structures of the dimerization domain of AraC determined from crystals grown in the absence and presence of arabinose show two prominent differences that could be the origin of the arabinose response (Soisson *et al.*, 1997). In the crystals grown in the absence of arabinose, each monomer of the protein interacts with two other monomers through two different interfaces, a face-to-face interaction between β-barrels as well as through a coiled-coil interface. This second interface is the only one found in the presence of arabinose. Conceivably then, the protein dimerizes by the face-to-face interaction in the absence of arabinose and by the coiled-coil interface in the presence of arabinose. Such a shift of dimerization interface would alter by 20 Å the distance between the points to which the DNA binding domains are attached. In turn, such a change in the separation of the DNA binding domains might be able to cause the protein to shift from preferring to loop to preferring to bind *cis*.

Alternatively, the arabinose-induced shift in AraC could result from movement of the N-terminal arm. This region of the protein dramatically changes conformation between the plus and minus arabinose states. In the absence of arabinose, the first 18 amino acids are disordered and are not visible in the electron density map, whereas, in the presence of arabinose, residues from the seventh on become visible as the arm folds over the bound arabinose and forms direct and indirect contacts with the sugar (Soisson *et al.*, 1997). The fact that the N-terminal arm of AraC is not visible in the absence of arabinose leaves open several possibilities. In this state, the arm may be disordered and non-functional, it may make specific contacts with the DNA-binding domain of AraC or the DNA, which were not present in the crystallization studies, or it may function in a non-specific way and merely occupy the space into which the DNA-binding domains would have to move in order that the protein be able to bind adjacent half-sites.

The accompanying paper (Seabold & Schleif, 1998) describes experiments indicating that the positioning of the DNA binding domains of AraC is more restricted in the absence of arabinose than in the presence of arabinose, a result more at odds with the alternative dimerization interfaces mechanism than in support of it. We have therefore carried out genetic experiments to assess the roles of the N-terminal arm and the DNA binding domain in the function of AraC. First, we studied the effects of deletions and point mutations in the first 20 amino acids of AraC. Since the carbonyl of Pro8 contacts arabinose directly and residues 9, 10, 12, and 13 make indirect contacts with arabinose, we expected to see a loss in the ability of arabinose to induce transcription activation when these residues were deleted or altered. Instead, such deletions made the protein constitutive in its transcription activating behavior at p_{BAD}. A variety of point mutations in the N-terminal arm made the protein constitutive, but none left p_{BAD} uninducible with a normal basal level. We also searched for mutations in the DNA binding domain of AraC protein that made the protein defective in transcription activation but not defective in DNA binding. Mutants were found with this apparent phenotype that were defective in their binding of arabinose.

Results

N-terminal arm deletions and point mutations

To determine whether the N-terminal arm of AraC is important to the regulatory properties of the protein, we mutated and deleted portions of the arm and measured induction of *ara* p_{BAD} and repression of p_C. For convenient assay of p_{BAD} activity, plasmids containing these mutant constructs were transformed into AraC$^-$ cells with a chromosomal copy of the promoter p_{BAD} of the *ara* regulatory region fused to the *lacZ* gene. Figure 2(a) shows the β-galactosidase levels in cells containing either wild-type AraC or various deletion mutants of AraC, in the presence and absence of arabinose. We found that deletion through Asn6 had little effect on p_{BAD} expression regulated by AraC, but that deletions through the seventh residue or beyond produced high constitutive transcription activation. Western blotting of cell extracts made from cultures grown in the absence of arabinose, (Figure 2(b)), shows that significant amounts of AraC are produced by the shorter deletions, indicating that the lack of constitutivity by the Δ6 mutant is not due to lack of AraC in the cells.

In view of the behavior of the deletion mutants, we sought to identify specific important locations within the first 20 amino acids of AraC. To do this we randomly mutagenized this region and screened for candidates that produced either con-

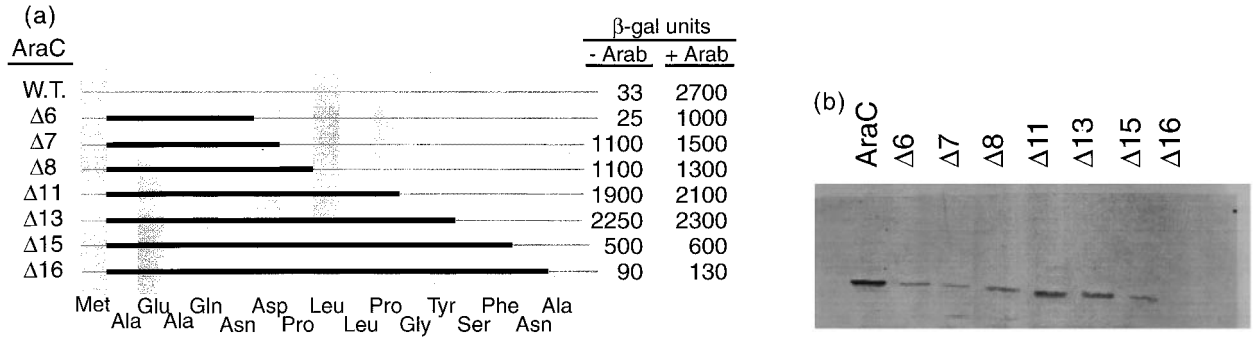

Figure 2. (a) Activity of p_{BAD} when stimulated by wild-type AraC or various N-terminal truncations in the presence and absence of arabinose as assayed by β-galactosidase levels generated from the p_{BAD}-*lacZ* fusion in strain RE5. The sequence of the N-terminal region is given and bars show the amino acids deleted. (b) Western blot of proteins extracted from the same aliquots of cells used for β-galactosidase measurements of cells grown in the absence of arabinose.

stitutive activation from p_{BAD} or possessed nearly normal basal levels and were unresponsive to arabinose. We shall call the latter type uninducible. Figure 3 shows the activity of p_{BAD} regulated by mutants found in the screen. While a number of point mutants produced constitutivity, no uninducible mutants were found. In a study of mutations conferring resistance to the anti-induction properties of D-fucose, Wallace (1982) found that most of his mutations (L9V, L9R, S14P, H18P, L19Q, V20G, G22C) were confined to the region now known to comprise the N-terminal arm, and most (all except L19Q) yielded constitutive acting AraC. The most commonly found and strongest acting mutations

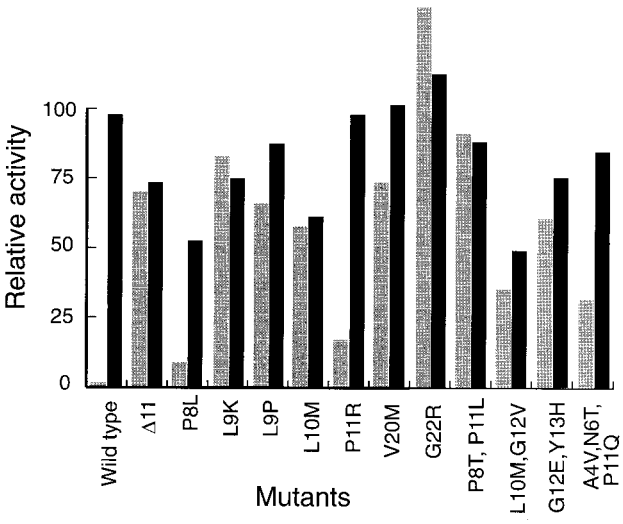

Figure 3. Activity of p_{BAD} in the presence and absence of arabinose when stimulated by wild-type AraC or various mutations in the N-terminal arm as assayed by β-galactosidase levels generated from the p_{BAD}-*lacZ* fusion in strain RE5. Mutant Δ11 is included for comparison. Light bars represent assays performed in the absence of arabinose, and dark bars represent assays performed in the presence of arabinose. Activity is graphed relative to that of fully activated wild-type AraC.

are those affecting leucine 9, which participates in only a single indirect contact with arabinose in the crystal structure of wild-type AraC.

Isolation of C-terminal domain mutations of AraC

We sought mutants specifically defective in induction of p_{BAD}, but not defective in DNA binding as assayed by repression of p_C. Because one common type would be mutants defective in arabinose binding to the N-terminal dimerization and arabinose binding domain, in hopes of finding other types of mutants, we specifically mutagenized the C-terminal region of AraC. This mutagenesis was done by PCR amplifying under mutagenic conditions the DNA coding for the C-terminal domain of AraC. The resulting DNA was ligated into an expression vector coding for the N terminus of AraC so as to regenerate a continuous AraC gene. The plasmid containing the regionally mutated AraC gene was then transformed into a strain that allowed convenient scoring of both the DNA binding ability of candidates by their ability to repress p_C-*lacZ*, and the inducing ability of candidates by their transcriptional activation of p_{BAD}-*araBAD*. 15,000 colonies were screened and 18 candidates were isolated of which seven carried multiple mutations. The remaining 11 candidates (Table 1) were characterized. *In vivo* activation by the candidates of p_{BAD} in the presence of arabinose ranged from one tenth of basal level, which means the protein actively represses or interferes with the promoter's basal activity, to a little more than twice basal level under conditions where wild-type AraC provided 31-fold induction. All the mutant proteins repressed p_C normally.

While it seems unlikely in light of *in vitro* DNA binding data to be presented below, it is conceivable that the mutant proteins fail to induce p_{BAD} because they no longer bind I_1. If this were so, they would have to bind O_1 to have passed the p_C repression screen for DNA binding. This possibility was eliminated by showing that the mutants

Table 1. Activation and repression properties of the C-terminal mutants

Number of isolates	Protein	Activation factor of p_{BAD}	Relative repressed level of p_C
–	Wild-type AraC	31	0.055
3	C183R	0.1	0.083
2	Q184R	1.3	0.040
1	D188G	1.0	0.043
1	Q230R	0.1	0.034
1	Q234R	2.2	0.040
1	K236R	2.0	0.050
1	N252Y	0.4	0.025
1	N252S	1.4	0.033

Transcription activation abilities at p_{BAD} and repression abilities at p_C of C-terminal mutants of AraC measured in strain SH288. Activation of p_{BAD}-*araBAD* was measured in the presence of 0.7 mM arabinose and is compared to the units per ml of arabinose isomerase present in the absence of AraC. Repression at p_C-*lacZ* was measured in the absence of arabinose and is compared to the level of β-galactosidase measured in the absence of AraC. The strain without AraC contained 54 arabinose units per cell and 330 units of β-galactosidase.

repress p_C in the presence or absence of arabinose in a construct in which O_1 has been inactivated so that looping between I_1 and O_2 is the only source of p_C repression (Table 2). Table 2 also shows that wild-type AraC represses *via* I_1-O_2 looping considerably better in the absence of arabinose than in the presence, thus demonstrating that repression of p_C in the absence of arabinose is *via* looping. In the looping assay, the mutant K236R did not repress p_C well upon the addition of arabinose. As a considerably higher arabinose concentration was used in this repression experiment than was used in the selection and scoring of induction deficient mutants, K236R is likely to interact with arabinose more weakly than wild-type AraC and/or to be a positive control type of mutant.

Weak arabinose response of the mutants

The AraC mutant proteins Q230R and N252 S were purified and tested for their abilities to stimulate open complex formation by RNA polymerase at p_{BAD}. Both mutant proteins activated like wild-type AraC (data not shown). Thus, the proteins possess the ability to activate transcription from p_{BAD}, but *in vivo* they fail to do so. A simple explanation for these properties is that the mutant proteins require higher than normal arabinose concentrations to shift them to the inducing state. Therefore we tested all the mutant proteins for their ability to respond to the sugar. It is difficult to assay the binding of arabinose to AraC in solution because the binding of arabinose is particularly weak, 0.1 mM; the protein binds irreversibly to many surfaces; the solubility of AraC is low; neither arabinose nor AraC show significant optical changes upon arabinose binding; and in our hands the fluorescence change of AraC upon arabinose binding is less than 4%. Therefore, to measure the affinity of arabinose binding to the mutant proteins, we used the fact that arabinose greatly reduces the dissociation rate of AraC from DNA. The dissociation in a fixed time interval was measured as a function of arabinose concentration using the DNA migration retardation assay. Figure 4(a) shows that in this assay wild-type protein needs between 0.05 and 0.5 mM arabinose to be stabilized in its DNA binding whereas AraC with a mutation in the arabinose binding pocket, Y82 S, requires between 5 and 100 mM arabinose for its binding to be stabilized. Figure 4(b) shows the altered arabinose responses of mutants Q230R and N252 S. Q230R AraC requires over 5 mM to be stabilized, while N252 S requires over 1.5 mM. The arabinose response of each mutant was tested and each required from 1.5 mM to over 5 mM of arabinose to be stabilized.

Table 2. Test of repression by looping

AraC protein	*lacZ* units from O_1^--p_C-*lacZ*	
	−Arabinose	+Arabinose
C^-	1652	–
Wild-type AraC	30	390
C183R	62	153
Q184R	14	27
D188G	56	125
Q230R	7	6
Q234R	12	23
K236R	84	346
N252Y	5	7
N252S	12	30

Expression in units of β-galactsidase from O_1^--p_C-*lacZ* in SH321 cells containing both the plasmid carrying the wild-type or mutant *araC* gene and the plasmid with the p_C-*lacZ* fusion. Cells were grown in the absence of arabinose and in the presence of 20 mM arabinose.

Suppressive effects of N-terminal arm mutations

The properties of our mutations can be understood if the changes interfere with positioning the N-terminal arm of AraC over arabinose when the sugar binds to the protein. Thus, if the mutations either create a binding site for the N-terminal arm on the C-terminal domain or strengthen binding to a pre-existing site on the C-terminal domain, then higher arabinose concentrations would be required to reposition the arm and generate the induction response. Six of the nine different induction deficient mutants we found result in a net positive charge change of the DNA binding domain. Possibly then, negatively charged amino acids in the N-terminal arm of AraC interact with the DNA binding domain through an electrostatic interaction. This is plausible as the only charged amino acids in the arm are both negatively charged. Table 3 shows that changing Asp7 to Ala7, D7A, in Q230R AraC results in a recovery of stimulation of p_{BAD} *in vivo*. Changing the other negatively charged amino acid in the N-terminal arm, Glu3, made the protein unstable *in vivo* and therefore its effect on Q230R AraC could not be determined.

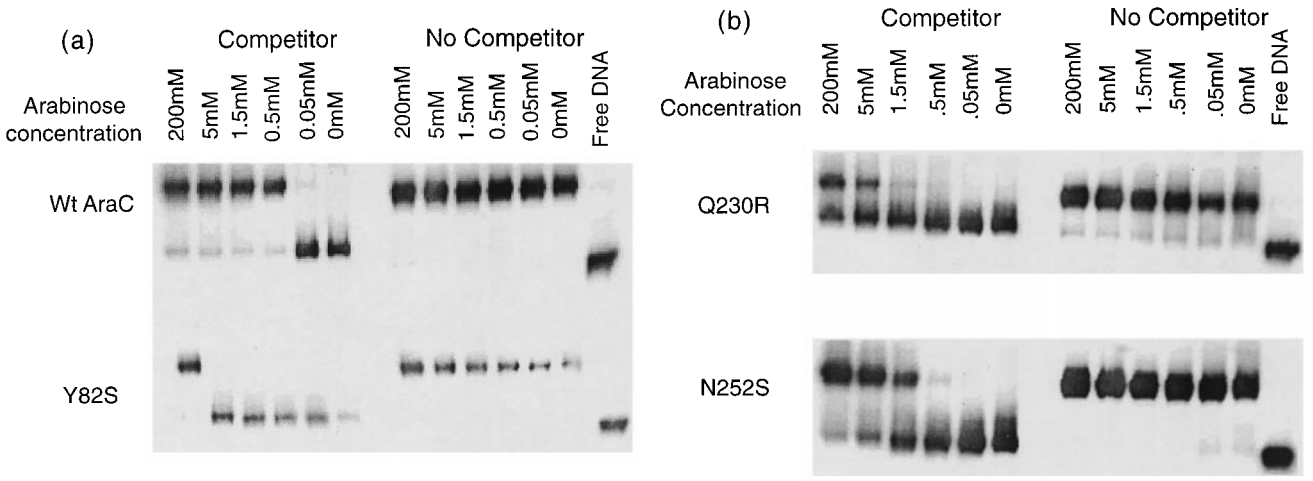

Figure 4. (a) Arabinose stabilization of AraC binding. I_1-I_2-p_{BAD} DNA was bound to wild-type or Y82S AraC at different arabinose concentrations so that just 100% of the DNA was bound. Samples were equilibrated for 20 minutes and then half the sample was loaded onto a 6% non-denaturing gel. Non-radioactive tight-binding competitor I_1-I_1-p_{BAD} DNA was added to the remainder of the ^{32}P-end-labeled samples and incubated ten minutes. The remaining sample was loaded onto the 6% non-denaturing polyacrylamide gel. A free DNA lane was also loaded for comparison. (b) Arabinose stabilization of proteins Q230R and N252S AraC.

The results are consistent with the idea that at least part of the interaction between the N-terminal arm of AraC and the DNA binding domain can be electrostatic. Not surprisingly, deletion of five N-terminal amino acids of AraC, Δ6, resulted in a partial recovery of induction activity at I_1-I_1-p_{BAD} and deletion of ten, Δ11, resulted in a complete recovery of the transcription activation activity of mutant Q230R (Table 4).

Discussion

In the first part of this work we found that deletions of the N-terminal arm of AraC removing residues two through seven or beyond, as well as alterations in the first 20 amino acids, some of which do and some of which do not contact arabinose, produce constitutive transcription activation behavior at p_{BAD}. We did not find any mutations in the N-terminal arm region that left p_{BAD} at a low level and would not activate p_{BAD} in the presence of arabinose. These unexpected findings indicate that the role of the N-terminal arm of AraC is more than simply helping arabinose bind to the dimerization domain. In the absence of arabinose, the arm apparently plays an active role in preventing the system from inducing p_{BAD}. Such a function is consistent with fact that the arm undergoes a major structural change upon the addition of arabinose (Soisson *et al.*, 1997). The N-terminal arm is unstructured in the absence of arabinose, but in the presence of arabinose residues seven and beyond fold over the sugar and become structured.

It seems plausible that random flailing about of the N-terminal arms, a so-called entropic brush mechanism, when not tied down by the presence of arabinose, excludes the DNA-binding domains from the vicinities of the arms and prevents the DNA binding domains from occupying positions necessary for binding to adjacent half-sites. While the brush mechanism could explain the intrinsic preference of AraC to loop (Seabold & Schleif, 1998), and the fact that deletions of the N-terminal arm make AraC constitutive, this mechanism leads to the prediction that the transition to constitutive activation will occur gradually as the arm is shortened, and not abruptly as was seen. Furthermore, the entropic brush mechanism does not explain the fact that AraC does not loop if the orientation of O_2 is reversed (Seabold & Schleif, 1998). Hence, we think it more likely that at least one critical amino

Table 3. Suppression of Q230R by the D7A mutation

AraC protein	Isomerase units from I_1-I_2-p_{BAD}-*araBAD*
Strain only	75
Wild-type AraC	5000
Q230R	56
D7A, Q230R	3700

Suppression of the defect in Q230R by D7A. Activation from I_1-I_2-p_{BAD}-*araBAD* in strain SH288 is given in units of arabinose isomerase per cell.

Table 4. Suppression of Q230R by N-terminal truncations

AraC protein	*lacZ* units from I_1-I_1-p_{BAD}-*lacZ*
Strain only	12
Wild-type AraC	500
Q230R	17
Δ6, Q230R	140
Δ11, Q230R	660

Suppression of the defect in Q230R by N-terminal truncations. Activation from I_1-I_1-p_{BAD}-*lacZ* in strain BS1 is given in units of β-galactosidase.

acid in the arm makes important interactions elsewhere in the absence of arabinose and holds the protein in its non-inducing state, a result also more in accord with the variety of point mutations in the N-terminal arm that lead to constitutive behavior at p_{BAD}.

We can infer from the facts described above that the N-terminal arm likely interacts with the DNA binding domain to make AraC prefer to loop. The data presented in the second part of this paper independently suggest the same. DNA coding for the C-terminal domain of AraC was mutagenized, transformed into AraC$^-$ cells, and candidates were chosen that were deficient in the induction of p_{BAD}, but normal in their repression of p_C. The candidates resulting from this screen proved to bind arabinose weaker than wild-type AraC. Since arabinose does not bind to the C-terminal domain of AraC, the effects of alterations there can most easily be understood as resulting from making it harder for the N-terminal arm of AraC to fold over arabinose when the sugar binds to the N-terminal domain. This could result from the creation or strengthening of sites in the DNA binding domain that bind the N-terminal arm in the absence of arabinose. Most of the mutations in the C-terminal domain increased its positive charge, and altering Asp7, one of the two charged amino acids in the N-terminal arm, to Ala suppressed the phenotype of the mutation. This is most easily understood to mean that in the mutants, at least part of the arm-domain interaction is electrostatic.

Overall then, our data suggest that in the absence of arabinose, the N-terminal arms of AraC interact with the C-terminal domains of the protein to hold them in an orientation that favors DNA looping. When arabinose is present, the N-terminal arms are pulled off the C-terminal domains and fold over the sugar that is bound to the N-terminal domain (Figure 5). None of the existing footprinting data, DNase I, dimethyl sulfate protection, premethylation interference, and hydrazine interference experiments (Carra & Schleif, 1993), revealed any differences in the details of AraC protein's contacting DNA in the presence and absence of arabinose. Thus, it seems unlikely, but certainly not excluded, that the N-terminal arm also contacts DNA in the absence of arabinose.

Although eight different mutations were found in the C-terminal domain affecting induction, several of them multiple times, all appear to reduce the affinity of arabinose for the protein, and none looks like a positive control type of mutation. Similar mutant screens performed on CAP protein and other transcription activators have yielded DNA binding-plus activation-defection mutations (Bell *et al.*, 1990; Eschenlauer & Reznikoff, 1991; Zhou *et al.*, 1993; Pratt & Silhavy, 1994; Gosink *et al.*, 1996; Whipple *et al.*, 1997). We suspect that our failure to find true positive control mutations in AraC results from the presence of redundant activation regions on the protein. Although, in principle, mutants with altered DNA sequence preferences would also have passed our screen, in view of their rarity in other proteins, we are not surprised at their absence here. In fact, the DNA binding abilities of the mutants in the presence and absence of arabinose were surprisingly similar to wild-type AraC as shown in the dissociation assays used to measure arabinose affinity. By virtue of the coupled equilibria governing the binding of arabinose and DNA to AraC, these results then permit us to conclude that the affinity of the mutant AraC proteins for arabinose is altered, not only in the presence of DNA, as was measured, but also in the absence of DNA. This follows since the ratio of affinities for arabinose in the presence and absence of DNA must equal the ratio of the affinities of AraC for DNA in the presence and absence of arabinose.

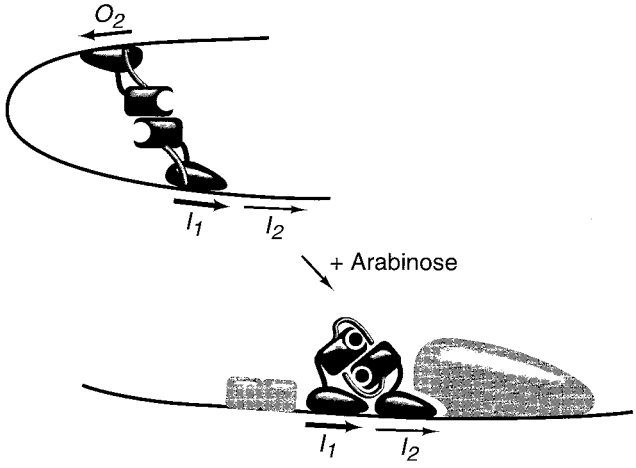

Figure 5. Proposed mechanism of the arabinose response of AraC. In the absence of arabinose the N-terminal arm of AraC binds to the DNA binding domain and holds it in a conformation that prefers to loop between the $araI_1$ and $araO_2$ half-sites. When arabinose binds, the N-terminal arm covers its binding site, freeing the DNA binding domains and allowing AraC to bind to adjacent half-sites.

It is not possible to infer much from the locations of the mutations in the C terminus. Even though they are scattered across seven different sites, it is possible that they are located close to one another within the tertiary structure of the protein. This cannot be known however, because the tertiary structure of neither the DNA binding domain of AraC nor that of any of the other members of the large family of proteins whose primary sequence is similar to that of the DNA binding domain of AraC (Gallegos *et al.*, 1997), has been determined. At present then, the only landmarks within the DNA binding domain are two regions with similarity to helix-turn-helix motifs found in some DNA binding proteins. The first of the potential helix-turn-helix regions, residues 197 to 216, within AraC could well adopt this structure as residues two and six of the potential recognition helix have been shown to make direct contacts with DNA

(Brunelle & Schleif, 1989). The second homology region, residues 246 to 265, may not contact DNA, as similar missing contact experiments failed to generate data supporting DNA contact for residues 1, 2, or 6 of the presumptive recognition helix (Brunelle & Schleif, 1989). Only one of the mutants, Asn252, lies in either of the two potential helix-turn-helix regions in AraC.

Three lines of evidence indicate that the constitutive behavior we observed in the N-terminal deletions and point mutants almost surely results from reduced or eliminated DNA looping by the mutant AraC. First, since looping has been shown to block access of RNA polymerase to p_{BAD} (Englesberg *et al*., 1969, in light of what was later learned about looping; Hahn *et al*., 1984) the high activity of p_{BAD} in the presence of our deletions or mutations in the N-terminal arm shows that looping by the constitutive mutants is absent or significantly reduced. Second, in a variant p_{BAD} promoter in which the positions of I_1 and I_2 have been interchanged and then moved so that I_1 overlaps the −35 region by two rather than four base-pairs, AraC activates only in the presence of arabinose (X. Zhang *et al*., 1996). Hence it follows that looping between the I_1 half-site of this construct and O_2 does not induce p_{BAD}, and thus it requires unusual suppositions to argue that the constitutive deletion mutants somehow can still be looping and also activate p_{BAD}. Third, in a plasmid in which O_1 has been inactivated, but p_C retains its activity, p_C is repressed by wild-type AraC looping between I_1 and O_2 in the absence of arabinose, but not in its presence. This shows that looping normally represses p_C in the absence of arabinose. The constitutive mutant Δ11 does not repress p_C in this construct (Seabold, 1997).

The fact that the N-terminal arm of AraC apparently plays a critical role in controlling the protein's behavior is not unprecedented. N and C-terminal arms on regulatory proteins or DNA binding proteins can play important roles in the proteins' interactions with DNA or other proteins. The N-terminal arm on lambda phage repressor, which is unstructured in solution, becomes structured and makes important contacts in the major grove of DNA when the protein binds DNA (Pabo *et al*., 1982). Many similar examples are known where the N-terminal arms of eukaryotic homeodomain proteins also contribute to specificity of DNA binding by becoming structured and binding to the DNA (e.g. Ades & Sauer, 1995). The role of arms on proteins is not limited, however, to contacting DNA. A number of examples are also known where the arm of a protein is unstructured until the protein forms a specific hetero-oligomer, mainly or entirely using contacts provided by the arm. These include interactions with homeodomains (Li *et al*., 1995), other eukaryotic transcription factors (H. Zhang *et al*., 1996; Xu *et al*., 1996), and signaling complexes (Keep *et al*., 1997). Interactions are also known between an arm and a different part of the same protein. In purified sigma-70 factor of the *E. coli* RNA polymerase, the N-terminal arm interferes with DNA binding by the C-terminal portion of the protein (Dombroski *et al*., 1993). Also, the C-terminal arm of RecA interferes with the protein's ability to bind to DNA and could interact directly with another part of RecA or with DNA (Tateishi *et al*., 1992).

In summary, genetic, physiological and biochemical data presented here and in the accompanying paper (Seabold & Schleif, 1998), indicate that in the absence of arabinose, the N-terminal arm of AraC protein interacts with the C-terminal DNA binding domain to hold it in a state where the protein prefers to engage in DNA looping interactions. When arabinose is added, folding of the arm over the sugar releases the DNA binding domain and, at the ara_{BAD} regulatory region, the protein then binds the adjacent I_1 and I_2 half-sites and induces transcription from p_{BAD}. This mechanism provides a versatile and general scheme for regulating the activity of a protein in response to the presence of a ligand. Part of the arm can possess an activating or inhibiting activity when it binds elsewhere on the protein until the presence of the ligand pulls it away. Possibly nature already widely uses such a scheme but it has not been frequently seen because crystallographers often remove unstructured parts of proteins with proteases to obtain crystals.

Materials and Methods

General methods

Arabinose isomerase was assayed as described, (Schleif & Wensink, 1981). β-Galactosidase assays were performed by the method of Miller (1972) as described by Maniatis *et al*. (1982). All assay results are the averages of at least duplicate measurements. For the N-terminal deletions and point mutations, strain RE5 (Δ*ara-leu1022* Δ*lac74 galK* str^r *thi1* [λ *ara*I_1-I_2-p_{BAD}-*lacZ*]; Eustance & Schleif, 1996) or strain SH321 (Δ*ara-leu1022* Δ*lac74 galK* str^r *thi1*; Hahn *et al*., 1984) was used, and for assay of p_C repression and p_{BAD} transcription activation of the C-terminal domain mutations, strain SH288 (F′*araC*$^-$ *araBAD*$^+$/Δ*ara-leu498* p_C-*lacZ* str^r Δ*lac*74 *thi1*; Hahn & Schleif, 1983) was used. Cell cultures were grown in M10 medium (Schleif & Wensink, 1981), 0.4% (w/v) Casamino acids, 10 μg/ml B1, 0.4% glycerol plus or minus 2% (w/v) at 37°C with shaking to an A_{550} between 0.3 and 0.9. Western blotting was performed as described (Eustance *et al*., 1994) on cells grown for β-galactosidase measurements as above. All plasmid constructs were made by standard molecular biological techniques (as described by Maniatis *et al*., 1982). All mutants were sequenced using double-stranded DNA sequencing (Kraft *et al*., 1988).

Isolation and characterization of AraC N-terminal domain deletion and point mutants

Deletions to AraC were made in AraC expression plasmid BB1, which is the same as plasmid pGB020 (Bustos & Schleif, 1993) except that the stop codon at the 3′ end of AraC is followed immediately by the *Sac*I restriction site sequence of the plasmid. Deletions 6, 11, and 16 were made using a single 3′ and various 5′ pri-

mers to extract the necessary AraC sequence from BB1 by polymerase chain reaction (PCR). The 3′ oligo had the sequence CAGCCAAGCTTAAGGTGCACGGC, and the 5′ oligos had the sequences: for Δ6, CATGCCATGGATCCCCTGCTGCCGCCATA; for Δ11, CATGCCATGGGATACTCGTTTAACGCCCAT; and for Δ16, CATGCCATGGCCCATCTGGTGGCGGGTTTA. The parental plasmid vector and PCR product inserts were then digested with *Nco*I and *Sac*I, the appropriate fragments were purified by gel electrophoresis, and ligated together. Point mutants were similarly made, except that the 5′ primer was synthesized such that each position underlined in the following sequence was doped with a mix of all nucleotides at one fiftieth the concentration of normal phosphoramidites. The sequence of the unaltered primer was CAGACCATGGCTGAAGCGCAAAATGATCCCCTGCTGCCGGGATACTCGTTTAACGCCCATCTGGTG. Other deletions were made by synthesizing DNA using *Pfu* DNA polymerase on one of the above templates, digesting away the parental plasmid DNA using *Dpn*I, then transforming into ultracompetent cells (Stratagene Quickchange® method). For Δ7, template was Δ6 and primer sequence was CACAGGAAACAGACCATGCCCCTGCTGCCGGG; for Δ8, template was Δ6 and primer sequence was CACAGGAAACAGACCATGCTGCTGCCGGGATAC; for Δ13, template was Δ11 and primer sequence was CACAGGAAACAGACCATGTCGTTTAACGCCCATCTG; and for Δ15, template was Δ11 and primer sequence was CACAGGAAACAGACCATGAACGCCCATCTGGTGGCG.

Isolation and characterization of C-terminal domain mutations

The AraC protein was cloned into the *Nco*I and *Xba*I sites of pSE380 (Invitrogen, San Diego) for overexpression of the protein *in vivo* (Bustos & Schleif, 1993). AraC mutants defective in the ability to activate transcription of the genes coding for the catabolic enzymes for arabinose were detected on tetrazolium arabinose plates. Reduced transcription from p_{BAD} will result in reduced catabolism of arabinose, yielding red colonies, whereas cells with wild-type transcription at p_{BAD} will appear white. Repression was monitored from the promoter p_C fused to the *lacZ* gene as AraC binding to p_C represses *lacZ* synthesis. Cells containing wild-type AraC plated on minimal salts, 0.4% glycerol, 0.4% (w/v) Casamino acids, 10 μg/ml B1 and 0.002% (x/v) X-gal, yield white colonies whereas cells containing AraC defective in DNA binding give blue colonies. Strain BS1 (F^- Δ*ara-leu1022* Δ*lac74 galK thi1* str^r [λ$araI_1$-I_1-p_{BAD}-*lacZ*]) was constructed from SH321 (Hahn *et al*., 1984) by the method of Simons *et al*. (1987). It contains λ phage in single copy on the chromosome carrying $araI_1$-I_1-p_{BAD}-*lacZ* with the P5 promoter (Reeder & Schleif, 1993).

A *Sal*I site was introduced into the DNA coding for the linker region of AraC changing bps 523 to 528 from ATGGAT to GTCGAC, pBS2. This created the conservative amino acid change M175V that resulted in a reduction in activation from p_{BAD} by a factor of 2 compared to wild-type AraC. Repression from p_C remained like wild-type. The DNA between the *Sal*I site and the *Xba*I site at the C terminus of AraC was amplified by PCR under mutagenic conditions (Leung *et al*., 1989). 5 ng of template DNA was added to 100 ng each of primer 1405, 5′-TCCATCCGCCAGTCGACAATCGGGTA-3′ and 1375, 5′-AAAACAGCCAAGCTTAAGGTGCACGG-3′. 10× reaction buffer (500 mM KCl, 200 mM Tris-HCl (pH 8.3), 15 mM $MgCl_2$, 5 mM $MnCl_2$) was added to a final 1× concentration in a total volume of 30 μl. 1 μl of *Taq* polymerase was added to each sample. Cycling parameters were 95°C one minute, 55°C one minute, and 72°C one minute. PCR products were purified on a 1.2% agarose gel, then cut with *Sal*I and *Xba*I overnight, cloned into pBS2, and transformed into strain SH288. Cells were plated on tetrazolium arabinose plates and screened for reduced utilization of arabinose. Candidate red colonies were patched onto minimal salts, 0.4 % glycerol, 0.4% Casamino acids, 10 μg/ml B1, and 0.002% X-gal plates. Colonies that are white indicate repression of p_C.

For the testing of repression of p_C by looping between I_1 and O_2, the *lacZ* gene, extracted from plasmid pTAP4 (Reeder & Schleif, 1993) by PCR using primers of sequence CCTCTAGACGGTTATTATTATTTTTG and TCTCCATGGAGGGAGTATGAAAAGTATGGTCGTTTTACAACGTCG, was inserted into plasmid pES51 between the *Nco*I site and an *Xba*I site that had been previously inserted into the middle of the pES51 *Aat*II site (making sequence ACGTCGCTCTAGAGCGACGT) to produce the plasmid p_C-*lacZ*. The promoter-occluding operator O_1 of plasmid p_C-*lacZ* was altered in seven locations such that O_1 is inactive, but p_C retains its activity. The final sequence for O_1 was TGAGCAAAGTGTCTCCGATCACGGTAGAAAAGTCCACA.

DNA migration retardation assay

The DNA migration retardation assay was performed with wild-type and mutant AraC proteins as described (Hendrickson & Schleif, 1984). Radiolabeled p_{BAD} DNA fragments were generated with PCR. 100 ng of ^{32}P-5′-end-labeled primer 5′-ATAATCACGGCAGAAAAGTCCA-3′ at 10^6 cpm/ng was mixed with 150 ng of unlabeled primer 5′-GTGCGCGTGCAGCCCTTATTGCCC-3′ and template plasmid pES51 containing the I_1-I_2-p_{BAD} promoter (Huo *et al*., 1988). PCR cycle parameters were 95°C one minute, 55°C one minute, 72°C one minute for 28 cycles. Crude cell lysates were prepared from cells over expressing the wild-type or mutant AraC proteins. Cells were grown to an A_{600} of 0.7 in YT broth (Schleif & Wensink, 1981). 3 ml of culture was centrifuged and resuspended in 0.5 ml 100 mM KPO_4 (pH 7.4), 50 mM KCl, 10% glycerol, 1 mM DTE, 0.1 mM $ZnCl_2$, 1 mM EDTA (pH 8). The resuspended cells were lysed by sonication and centrifuged at 8500 *g* for ten minutes. The supernatant was removed and 170 μl of 100% glycerol was added to 500 μl of supernatant. The lysates were then stored at −70°C for up to two weeks. Binding reactions were carried out in 10 mM Tris-OAC (pH 7.4), 1 mM EDTA, 50 mM KCl, 1 mM DTT, 5% glycerol, 50 ng calf thymus DNA/μl. Protein from the lysates was added so that just 100% of 1 ng of I_1-I_2 ^{32}P-end-labeled DNA was bound. Binding reactions were equilibrated for 20 minutes and half the sample was loaded onto a non-denaturing 6% acrylamide, 0.1% MBA gel. A 100x molar excess of non-radioactive specific competitor DNA was added to each sample, and after ten minutes, the remainder was loaded on the non-denaturing gel.

Acknowledgments

We are grateful to the members of this laboratory for critical discussions and to Ula Gryczynska for technical

assistance. This material is based upon work supported by the National Institutes of Health grant GM18277 to R.F.S.

References

Ades, S. E. & Sauer, R. T. (1995). Specificity of minor-groove and major-groove interactions in a homeodomain-DNA complex. *Biochemistry,* **34**, 14601–14608.

Bell, A., Gaston, K., Williams, R., Chapman, K., Killes, A., Buc, H., Minchin, S., Williams, J. & Busby, S. (1990). Mutations that alter the ability of the *Escherichia coli* cyclic AMP receptor protein to activate transcription. *Nucl. Acids Res.* **18**, 7243–7250.

Brunelle, A. & Schleif, R. (1989). Determining residue-base interactions between AraC Protein and *araI* DNA. *J. Mol. Biol.* **209**, 607–622.

Bustos, S. A. & Schleif, R. F. (1993). Functional domains of the AraC protein. *Proc. Natl Acad. Sci. USA,* **90**, 5638–5642.

Carra, J. & Schleif, R. (1993). Variation of half-site organization and DNA looping by AraC protein. *EMBO J.* **12**, 35–44.

Dombroski, A. J., Walter, W. A. & Gross, C. A. (1993). Amino-terminal amino acids modulate sigma-factor DNA-binding activity. *Genes Dev.* **7**, 2446–2455.

Dunn, T., Hahn, S., Ogden, S. & Schleif, R. (1984). An operator at −280 base pairs that is required for repression of *araBAD* operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc. Natl Acad. Sci. USA,* **81**, 5017–5020.

Englesberg, E., Squires, C. & Meronk, F., Jr (1969). The L-arabinose operon in *Escherichia coli* a genetic demonstration of two functional states of the product of a regulator gene. *Proc. Natl Acad. Sci. USA,* **62**, 1100–1107.

Eschenlauer, A. C. & Reznikoff, W. S. (1991). *Escherichia coli* catabilite gene activator protein mutants defective in positive control of *lac* operon transcription. *J. Bacteriol.* **173**, 5024–5029.

Eustance, R. J. & Schleif, R. F. (1996). The linker region of AraC protein. *J. Bacteriol.* **178**, 7025–7030.

Eustance, R. J., Bustos, S. A. & Schleif, R. F. (1994). Reaching out: locating and lengthening the interdomain linker in AraC protein. *J. Mol. Biol.* **242**, 330–338.

Gallegos, M., Schleif, R., Bairoch, A., Hofmann, K. & Ramos, J. (1997). The AraC/XylS family of transcriptional regulators. *Microbiol. Mol. Biol. Rev.* **61**, 393–410.

Gosink, K. K., Gaal, T., Bokal, A. J. & Gourse, R. L. (1996). A positive control mutant of the transcription activator protein FIS. *J. Bacteriol.* **178**, 5182–5187.

Greenblatt, J. & Schleif, R. F. (1971). Arabinose C protein regulation of the arabinose operon *in vitro*. *Nature New Biol.* **233**, 166–170.

Hahn, S. & Schleif, R. F. (1983). *In vivo* regulation of the *Escherichi coli araC* promoter. *J. Bacteriol.* **155**, 593–600.

Hahn, S., Dunn, T. & Schleif, R. F. (1984). Upstream repression and CRP stimulation of the *Escherichia coli* L-arabinose operon. *J. Mol. Biol.* **180**, 61–72.

Hendrickson, W. & Schleif, R. F. (1984). Regulation of the *Escherichia coli* L-arabinose operon studied by gel electrophoresis DNA binding assay. *J. Mol. Biol.* **178**, 611–628.

Hendrickson, W. & Schleif, R. (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.

Huo, L., Martin, K. & Schleif, R. F. (1988). Alternate DNA loops regulate the arabinose operon in *Escherichia coli*. *Proc. Natl Acad. Sci. USA,* **85**, 5444–5448.

Keep, N. H., Barnes, M., Barsukov, I., Badii, R., Lian, L., Segal, A., Moody, P. C. E. & Roberts, G. C. K. (1997). A molulator of rho family G protein, rhoGDI, binds these G proteins via an immunoglobulin-like domain and a flexible N-terminal arm. *Structure,* **5**, 623–633.

Kraft, R., Tardiff, J., Krauter, K. S. & Leinwand, L. A. (1988). Using mini-prep plasmid DNA for sequencing double stranded templates with Sequenase. *BioTechniques,* **6**, 544–546.

Lee, D. & Schleif, R. F. (1989). *In vivo* DNA loops in *araCBAD*: size limits and helical repeat. *Proc. Natl Acad. Sci. USA,* **86**, 476.

Lee, N., Wilcox, G., Gielow, W., Arnold, J., Cleary, P. & Englesberg, E. (1974). *In vitro* activation of the transcription of *araBAD* operon by *araC* activator. *Proc. Natl Acad. Sci. USA,* **71**, 634–638.

Lee, N., Francklyn, C. & Hamilton, E. P. (1987). Arabinose-induced binding of AraC protein to $araI_2$ activates the *araBAD* operon promoter. *Proc. Natl Acad. Sci. USA,* **84**, 8814–8818.

Leung, D. W., Chen, E. & Goeddel, D. V. (1989). A method for random mutagenesis of a defined DNA segment using modified polymerase chain reaction. *Technique,* **1**, 11–165.

Li, T., Stark, M. R., Johnson, A. D. & Wolberger, C. (1995). Crystal structure of the MATa1/MATα2 homeodomain heterodimer bound to DNA. *Science,* **270**, 262–269.

Lobell, R. & Schleif, R. F. (1990). DNA looping and unlooping by AraC protein. *Science,* **250**, 528–532.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Martin, K., Hou, L. & Schleif, R. F. (1986). The DNA loop model for *ara* repression: AraC protein occupies the proposed loop sites *in vivo* and repression-negative mutations lie within these same sites. *Proc. Natl Acad. Sci. USA,* **83**, 3654–3658.

Miller, J. H. (1972). *Experiments in Molecular Genetics,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Pabo, C. O., Krovatin, W., Jeffrey, A. & Sauer, R. T. (1982). The N-terminal arms of lambda repressor wrap around the operator DNA. *Nature,* **298**, 441–443.

Pratt, L. A. & Silhavy, T. J. (1994). OmpR mutants specifically defective for transcriptional activation. *J. Mol. Biol.* **243**, 579–594.

Reeder, T. & Schleif, R. (1993). AraC protein can activate transcription from only one position and when pointed in only one direction. *J. Mol. Biol.* **231**, 205.

Schleif, R. F. & Wesink, P. (1981). *Practical Methods in Molecular Biology,* Springer-Verlag, New York.

Seabold, R. (1997). PhD thesis, Biology Department, Johns Hopkins University.

Seabold, R. & Schleif, R. (1998). Apo-AraC actively seeks to loop. *J. Mol. Biol.* **278**, 529–538.

Simons, R. W., Houman, F. & Kleckner, N. (1987). Improved single and multicopy *lac*-based cloning

vectors for protein and operon fusions. *Gene.* **53**, 85–96.
Soisson, S. M., MacDougal-Shackleton, B., Schleif, R. F. & Wolberger, C. (1997). Structural basis for ligand-regulated oligomerization of AraC. *Science,* **276**, 421–425.
Tateishi, S., Horii, T., Ogawa, T. & Ogawa, H. (1992). C-terminal truncated *Escherichia coli* RecA protein RecA5327 has enhanced binding affinities to single and double-stranded DNAs. *J. Mol. Biol.* **223**, 115–129.
Wallace, R. (1982). PhD thesis, University of California, Santa Barbara.
Whipple, F. W., Ptashne, M. & Hochschild, A. (1997). The activation defect of a lambda cI positive control mutant. *J. Mol. Biol.* **265**, 261–265.
Wilcox, G. & Meuris, P. (1976). Stabilization and size of AraC. *Mol. Gen. Genet.* **145**, 97–100.
Xu, J., Nawaz, Z., Tsai, S., Tsai, M. & O'Malley, B. (1996). The extreme C terminus of progesterone receptor contains a transcriptional repressor domain that functions through a putative corepressor. *Proc. Natl Acad. Sci. USA,* **93**, 12195–12199.
Zhang, H., Catron, K. M. & Abate-Shen, C. (1996). A role for the Msa-1 homeodomain in transcriptional regulation: residues in the N-terminal arm mediate TATA binding protein interaction and transcriptional repression. *Proc. Natl Acad. Sci. USA,* **93**, 1764–1769.
Zhang, X. & Schleif, R. (1998). CAP mutations affecting activity of the *araBAD* promoter. *J. Bacteriol.* **180**, 195–200.
Zhang, X., Reeder, T. & Schleif, R. F. (1996). Transcription activation parameters at *ara* p_{BAD}. *J. Mol. Biol.* **258**, 14–24.
Zhou, Y., Zhang, X. & Ebright, R. H. (1993). Identification of the activating region of catabolite gene activator protein (CAP): isolation and characterization of mutants of CAP specifically defective in transcription activation. *Proc. Natl Acad. Sci. USA,* **90**, 6081–6085.

Edited by M. Yaniv

(Received 22 September 1997; received in revised form 21 January 1998; accepted 5 February 1998)