J. Mol. Biol. (1998) **278**, 529–538

JMB

Apo-AraC Actively Seeks to Loop

Robert R. Seabold and Robert F. Schleif*

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

In the absence of arabinose and interactions with other proteins, AraC, the activator-repressor that regulates the *araBAD* operon in *Escherichia coli*, was found to prefer participating in DNA looping interactions between the two well-separated DNA half-sites, $araI_1$ and $araO_2$ at their normal separation of 211 base-pairs rather than binding to these same two half-sites when they are adjacent to one another. On the addition of arabinose, AraC preferred to bind to the adjacently located half-sites. Inverting the distally located $araO_2$ half-site eliminated the looping preference. These results demonstrate that apo-AraC possesses an intrinsic looping preference that is eliminated by the presence of arabinose. We developed a method for the accurate determination of the relative affinities of AraC for the DNA half-sites $araI_1$, $araI_2$, and $araO_2$ and non-specific DNA. These affinities allowed accurate calculation of basal level and induced levels of expression from p_{BAD} under a wide variety of natural and mutant conditions. The calculations independently predicted the looping preference of apo-AraC.

Keywords: AraC; DNA looping; relative affinities; gene expression; partition function

*Corresponding author

Introduction

In the absence of arabinose, the dimeric regulator of the *ara* p_{BAD} promoter, AraC, forms a DNA loop between the I_1 and O_2 half-sites (Figure 1) that represses activity of p_{BAD} (Dunn *et al.*, 1984; Hahn *et al.*, 1984; Martin *et al.*, 1986; Huo *et al.*, 1988; Lee & Schleif, 1989). When arabinose is added, looping is reduced and AraC primarily binds to the two adjacent half-sites, I_1 and I_2, where it activates transcription from p_{BAD} (Greenblatt & Schleif, 1971; Hendrickson & Schleif, 1984, 1985; Lee *et al.*, 1987; Lobell & Schleif, 1990; for a review see Schleif, 1992). The O_1 pair of half-sites, whose occupancy represses the promoter for the synthesis of AraC protein, p_C, does not appear to play a direct role in the regulation of p_{BAD}.

Because AraC can bind to two half-sites in direct or inverted repeat orientation, or when the half-sites have been separated by an additional ten base-pairs (Carra & Schleif, 1993), attractive mechanisms for AraC action include free rotation of the protein's two DNA binding domains. A general mechanism based on such freedom allows the protein to distribute itself amongst the various states available to the system in accordance with the free energy differences between the states. Such a mechanism could include the effect of arabinose either through a sugar-induced change in the protein's relative affinities for the various half-sites involved, or a sugar-induced restriction in the freedom of the DNA binding domains. The latter has seemed more likely as AraC, which consists of functionally independent dimerization-arabinose binding and DNA binding-transcription activation domains (Bustos & Schleif, 1993), has been postulated on the basis of the crystal structures in the presence and absence of arabinose and aggregation properties (Soisson *et al.*, 1997) to change the separation between the attachment points for the DNA binding domains through an arabinose regulated change in the dimerization interface.

In the work described here we have measured the relative affinity of AraC protein for the half-sites relevant to p_{BAD} expression, I_1, I_2, O_2, and non-specific DNA in the presence and absence of arabinose. As expected, arabinose did not significantly change the relative affinities for these half-sites. Contrary to expectations, however, AraC was found in the absence of arabinose to possess an intrinsic preference to loop. This was revealed by direct *in vivo* footprinting measurements in which the I_2 half-site at p_{BAD} was converted to the O_2 sequence, and thus AraC had the opportunity to loop between I_1 and either the distal O_2 half-site or bind I_1 and the adjacent half-site. Additionally, and again not as expected from the alternative dimerization interfaces mechanism, AraC did not prefer to loop when the orientation of the distal O_2 half-

0022–2836/98/180529–10 $25.00/0/mb981713

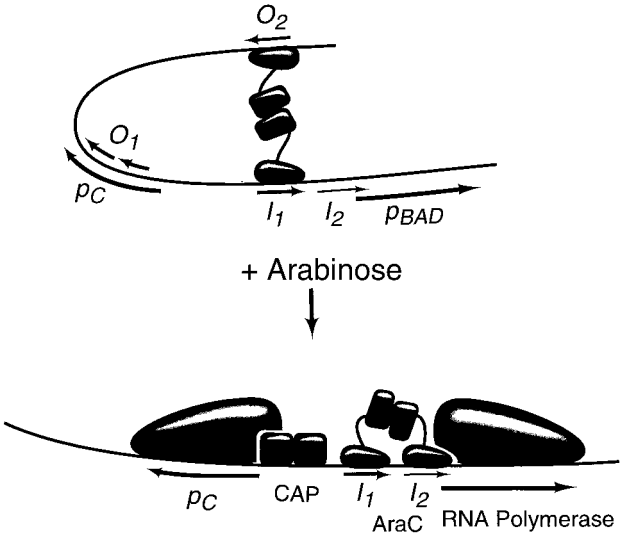


Figure 1. The regulatory region of the *araCBAD* operon. AraC, in the absence of arabinose, loops out the intervening DNA between I_1 and O_2 preventing induction, and in the presence of arabinose, AraC binds the adjacent I_1 and I_2 half-sites and activates p_{BAD}. CAP stimulates both p_{BAD} and p_C.

site was reversed. The preference of AraC to loop was also revealed in calculations of the expression levels of p_{BAD} under a wide variety conditions. These were made possible from the determination of the relative binding affinities for the various half-sites in the presence and absence of arabinose.

Results and Discussion

Measurement of relative affinities of AraC for I_1, O_2, I_2 and non-specific DNA

To overcome the problems inherent in measuring absolute binding affinities, we accurately measured the relative affinities for AraC binding to various sites and formulated a quantitative description that utilized these relative affinities.

The relative affinities of a protein for two DNA binding sites may be obtained without knowledge of the actual solution free protein concentration as follows. The equilibrium affinity constant describing binding of a protein to its operator may be written:

$$K_a = \frac{[P\text{-}O]}{[P] \times [O]}$$

where [P] is the concentration of the protein, [O] is the concentration of the free operator, and [P-O] is the concentration of the protein-operator complex. For two non-competing binding reactions carried out in the same reaction vessel, with protein binding to DNA species one, containing operator O_1, and to DNA species two, containing operator O_2, a combined equation may be written:

$$\frac{K_{a1}}{K_{a2}} = \frac{\dfrac{[P - O_1]}{[P] \times [O_1]}}{\dfrac{[P - O_2]}{[P] \times [O_2]}}$$

The free protein concentration cancels out from the above equation, leaving the ratio of the K_as equal to the ratio of the ratios of bound to free DNA for each of the operators. These two ratios are readily quantified in a gel-based DNA migration retardation assay if the four subspecies of DNA: O_1, P-O_1, O_2, and P-O_2, are well separated from one another after electrophoresis.

To measure the relative affinities of AraC for the three half-sites I_1, I_2, and O_2, and for non-specific DNA, we constructed a set of DNA fragments containing the sites I_1-I_1, I_1-O_2, I_1-I_2, and I_1-non-half-site DNA (Figure 2). With one half-site identical in each member of the set and the immediate context of the operator maintained identically, differences in affinity measured between two operators are due only to the difference of affinity of AraC for the second half-site. The operators were compared

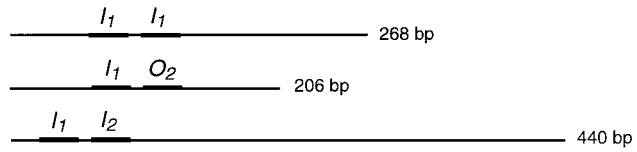


Figure 2. Artificial operator constructs. The length of fragments and operator position relative to the fragments' ends vary to allow separation of both free and AraC-bound bands by electrophoresis. ns, non-half-site DNA.

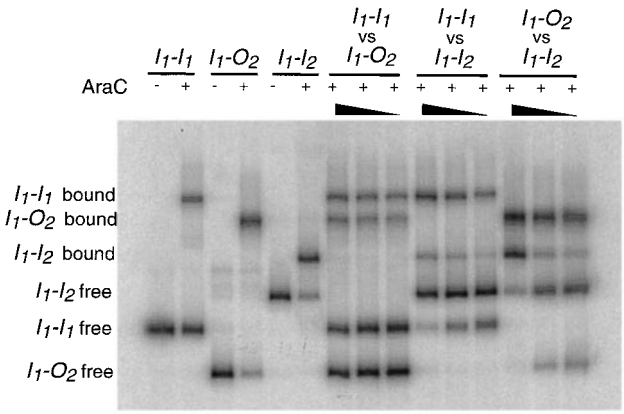


Figure 3. DNA migration retardation assay performed in the presence of arabinose. Three two-half-site operators are compared against each other. Each pair of operators was assayed at three different AraC concentrations. Although AraC bends DNA, the sample containing the I_1-I_2 site was the least retarded of the DNAs because the AraC binding site was close to the end.

Table 1. Ratios of DNA binding affinities

Sites present on the DNA	−Arabinose	+Arabinose
I_1-I_2 *vs* I_1-I_2	170 ± 6	70 ± 10
I_1-I_1 *vs* I_1-O_2	10 ± 5	3 ± 0.5
I_1-O_2 *vs* I_1-I_2	18 ± 2	20 ± 10
I_1-I_2 *vs* I_1-*ns*	4 ± 1	10 ± 5

pair-wise by DNA migration retardation assay (Figure 3), and the complete results of the DNA migration retardation assays are shown in Table 1.

Because of their loose connections through the dimerization domains (Bustos & Schleif, 1993; Carra & Schleif, 1993; Eustance *et al.*, 1994; Eustance & Schleif, 1996), it seems highly unlikely that the DNA sequence to which one DNA binding domain of the dimeric AraC protein binds affects the sequence selectivity of the other DNA-binding domain. This assumption can be verified experimentally by ascertaining whether, for binding amongst a group of sites A, B, and C:

$$\frac{\frac{K_A}{K_B}}{\frac{K_A}{K_C}} = \frac{K_C}{K_B}$$

Experimentally, we found:

$$\frac{K_{I1I1}}{K_{I1I2}} = 170,\ \frac{K_{I1I1}}{K_{I102}} = 10,\ \frac{K_{I102}}{K_{I1I2}} = 18$$

Since 170/10 is approximately 18, the data show good internal consistency. Internal consistency is also well satisfied in the absence of arabinose, thus validating the method and the assumption of binding site independence.

AraC possesses an intrinsic DNA-looping preference

We also tested whether AraC possesses an intrinsic preference to loop between distally located half-sites rather than bind to adjacent half-sites. This was done by replacing the I_2 half-site with the O_2 half-site in the *araBAD* regulatory region (Figure 4(a)). The remainder of the partially overlapping RNA polymerase binding site at p_{BAD} was not present so as to avoid any potential cooperative interactions between AraC and RNA polymerase in the unlikely event that the change from I_2 to O_2 did not itself inactivate p_{BAD}. In this construct, both looping and *cis* interactions of AraC involve binding to one I_1 half-site and to one O_2

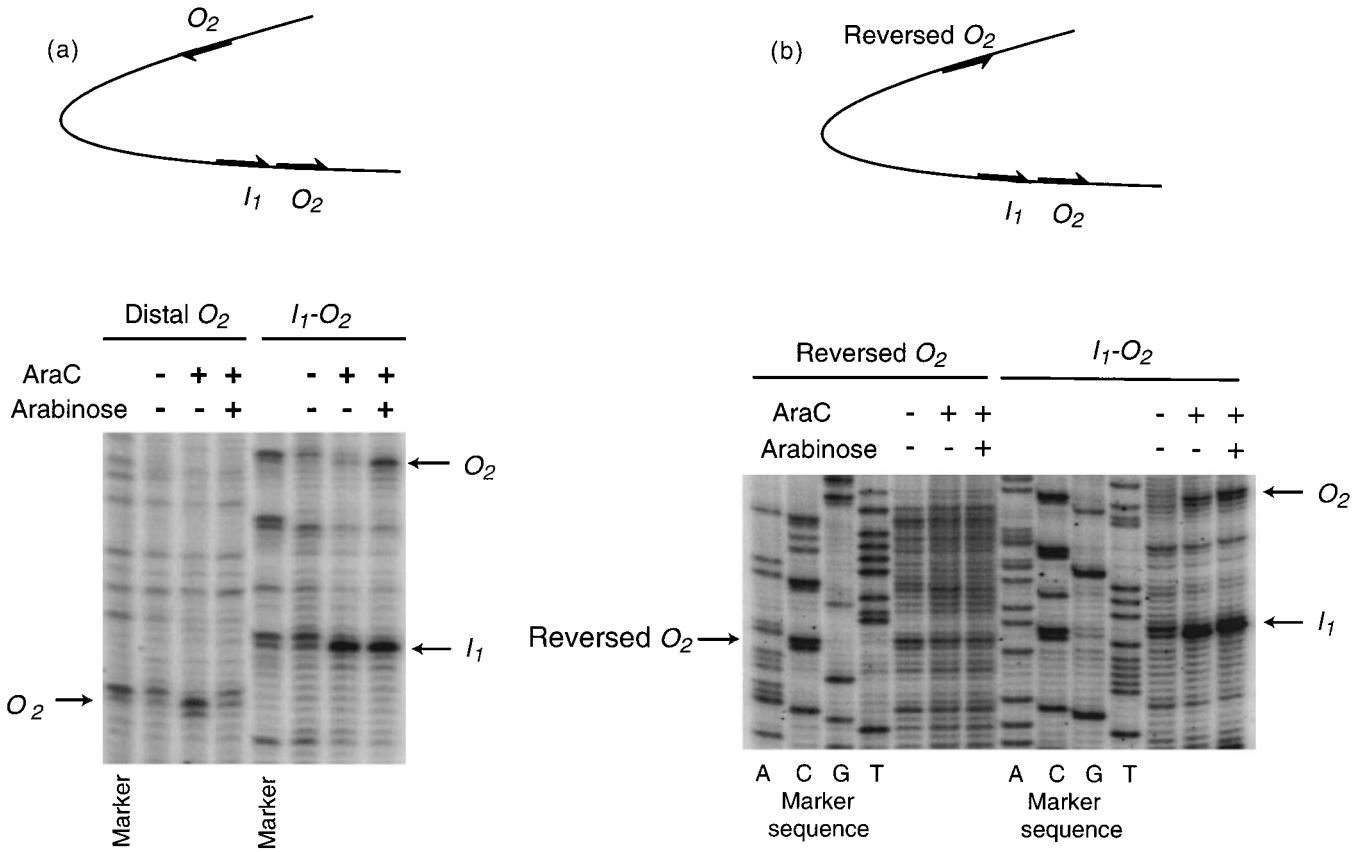


Figure 4. (a) Experimental construct used for DMS *in vivo* footprint assay showing O_2 in place of downstream half-site I_2 and the results of DMS *in vivo* footprinting this DNA. Marker is the Cs lane from a sequencing reaction and represents all possible complementary Gs that might be labeled by DMS. (b) Experimental construct used for DMS *in vivo* footprint assay showing reversed distal O_2 half-site and the results of DMS *in vivo* footprinting this DNA. Arrows mark the positions of the bands that reflect occupancy of the noted half-sites.

half-site. Thus, the partitioning of AraC in this system between the two states should reflect its intrinsic preference or reluctance to loop. This partitioning was measured with *in vivo* footprinting. Occupancy of the O_2 half-site adjacent to I_1 indicates binding *cis*, and occupancy of the distal half-site indicates looping. As reflected in the relative occupancies of the two half-sites, in the absence of arabinose, (Figure 4(a)), AraC does indeed prefer to loop, and in the presence of arabinose, it prefers to bind *cis*.

Apparently, the two DNA binding domains of the dimeric AraC protein are held such that the protein prefers to engage in DNA looping interactions in the absence of arabinose. If that is indeed the case, reversing the direction of the distally located O_2 half-site while maintaining the major groove regions that are contacted by AraC on the same face of the DNA as before should interfere with DNA looping. The I_1-O_2 portion of the experimental plasmid thus remained the same as for the plasmid described above. Figure 4(b) shows that very little looping occurs, with arabinose present or absent, in the construct with O_2 reversed.

Quantitative description of the *ara* regulatory system

Although cells grow and divide, the time scales relevant to the binding of AraC protein to its regulatory sites are so short that the binding of AraC can be considered to be near equilibrium. Thus, AraC could merely partition itself amongst the various binding states of the system (Figure 5), in accordance with the free energy differences amongst them. Consequently, we should be able to derive a quantitative description that predicts, under a wide variety of conditions, the fraction of operon copies in which AraC loops between I_1 and O_2 to repress p_{BAD} and the fraction in which AraC binds to I_1-I_2 to induce p_{BAD}. Such a calculation requires accurate knowledge of the binding affinities of AraC to the various sites involved, as was determined above. Since expression of p_{BAD} appears to be proportional to the amount of the DNA that has AraC bound at I_1 and I_2 (Zhang *et al.*, 1996), the experimentally measured expression levels of the *araBAD* operon can be compared to the calculations.

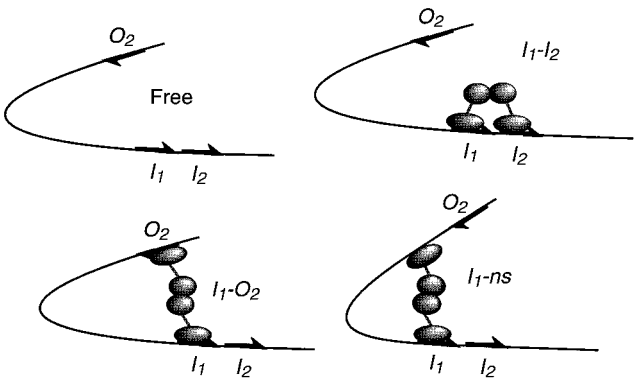


Figure 5. The four relevant states for which occupancies were calculated: no AraC bound; AraC bound *cis* to half-sites I_1 and I_2; AraC bound to half-sites I_1 and O_2 and forming a DNA loop; and AraC bound to half-site I_1 and non-specific, *ns*, DNA.

Ackers and co-workers successfully used a thermodynamic approach like that outlined above to calculate the behavior of the phage λ repressor-O_R system (Ackers *et al.*, 1982). The method has also been successfully applied to the *lac* (Brenowitz *et al.*, 1991; Senear & Brenowitz, 1991) and *gal* operon repressor systems (Brenowitz *et al.*, 1990; Senear & Brenowitz, 1991; Dalma-Weiszhausz & Brenowitz, 1992). Although the phage, *lac*, and *gal* systems are made up of multiple operator binding sites, as is the *ara* system, they differ from *ara* in that AraC is affected by a ligand, and in the *ara* system the operators are not adjacent to one another.

Three important parameters needed in the calculation of the *ara* system regulatory behavior could not easily be determined by experiment. One of these is the conversion of the relative affinities for two independent half-sites to the affinity for AraC binding to the two half-sites simultaneously. The other two, one for the presence and one for absence of arabinose, provide for the overall preference or reluctance of AraC to engage in a DNA looping reaction with half-sites separated by several hundred base-pairs rather than engage in a binding reaction with adjacent half-sites. This overall preference is a combination of any intrinsic looping preference and any effects due to AraC-RNA polymerase interactions. These parameters were therefore included in the equations as unknowns. As described in the Appendix, the values of these three unknowns were determined numerically by requiring that the equations correctly predict the p_{BAD} uninduced basal level, its fully induced level, and its basal level when the O_2 half-site is absent. The equations were solvable and the solutions were not highly sensitive to the precise values of the known or unknown parameters.

It is striking that the simplest description of the *ara* p_{BAD} regulatory system consistent with the known biochemical details accommodates the system's essential regulatory features. As a further check of the adequacy of the description, we tested the calculations on additional regulatory region variants (Figure 6). For three of these constructs we replaced the upstream O_2 half-site with the I_1, or I_2, half-sites, or non-half-site DNA and measured the levels of expression of the reporter β-galactosidase from p_{BAD} in the basal and induced states. The basal and induced levels fell by the predicted amounts when O_2 was replaced with the tighter-binding half-site I_1, and rose by the predicted amounts when O_2 was replaced with the weaker-binding half-site I_2 and non-half-site DNA (Figure 7(a) and (b)). Earlier observations that are also predicted by the quantitative description

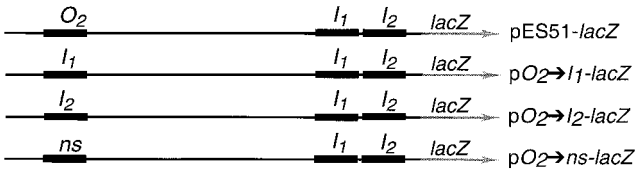
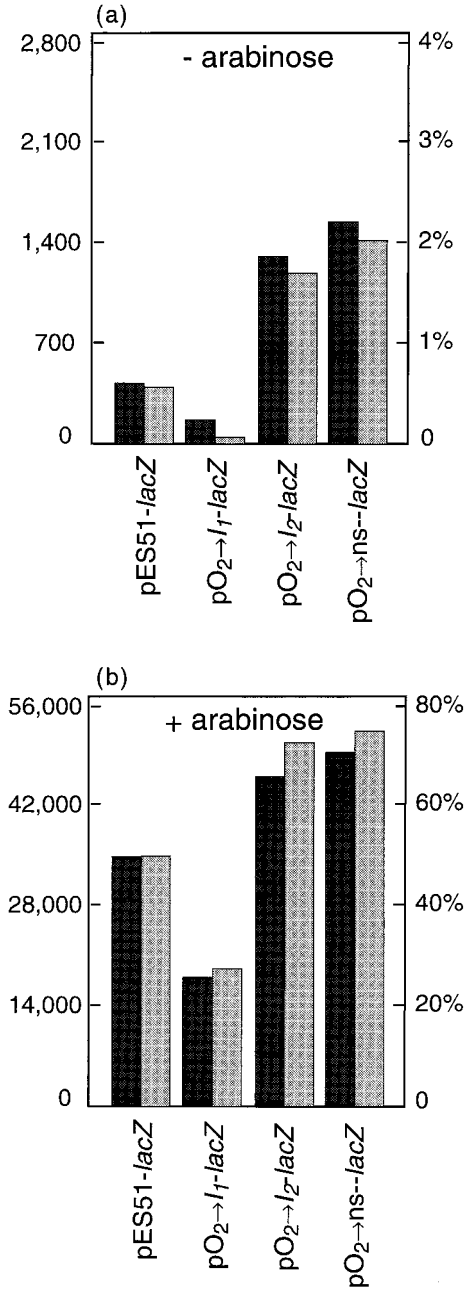


Figure 6. Structures of the *ara* upstream regulatory regions of the constructs used to verify the accuracy of predictions made using our descriptive equations.

include: the slight increase in induction level when AraC concentration is increased 50-fold (Hahn *et al.*, 1984); and the fact that replacing I_2 with I_1(Reeder & Schleif, 1993) makes p_{BAD} constitutive.



The values determined for the unknown parameters reveal that, in the absence of arabinose, AraC possesses a substantial preference to loop between two well-separated half-sites rather than bind to two adjacent half-sites. The importance of the parameter giving the preference to loop is shown in two ways. If it is set to one in our calculations while leaving the other parameters unaltered, the quantitative calculations became highly inaccurate and predict greatly elevated basal levels. Alternatively, if the looping preference is fixed at unity, the resulting set of equations cannot be solved as described above and in the Appendix, as no values for the remaining parameters will solve the equations. Figure 7(c) compares the measured and predicted expression levels for the above-mentioned constructs if the looping preference is set equal to one. These results further indicate the importance of the preference to loop in making accurate predictions.

Historically, we used the relative half-site affinities to calculate expression of p_{BAD} under various conditions. While the calculations accurately predicted many previously known results as well as new results, they also made the entirely unexpected prediction that AraC possesses an intrinsic

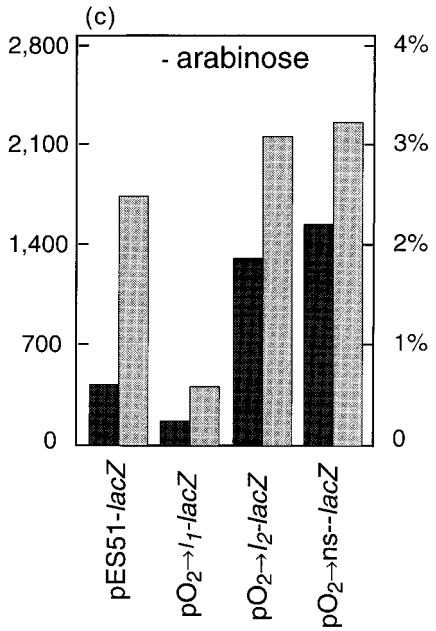


Figure 7. (a) Comparison of the results of β-galactosidase assays with predictions of the occupancy of half-site I_2 in the absence of arabinose. Dark bars represent assayed experimental data and light bars represent predictions. Expression from p_{BAD} is assumed to be directly proportional to occupancy of I_2. β-Galactosidase levels are expressed in Miller units and predictions as % occupancy of I_2. (b) Same as (a) but in the presence of arabinose. (c) Comparison of data from β-galactosidase assays with predictions made when the parameter for the preference to loop is set to one. Dark bars represent assayed expression data and light bars represent predictions.

preference to loop between distantly located half-sites. We then checked the prediction of a preference to loop with *in vivo* footprinting in a construct in which AraC had its choice of either looping between well separated I_1 and O_2 half-sites or binding *cis* to adjacently located I_1 and O_2 half-sites. Not only did AraC prefer to loop, inverting the distal O_2 half-site eliminated the looping preference. As discussed below, these results indicate that the orientations of the DNA binding domains of AraC are constrained in the absence of arabinose.

Concluding discussion

This paper presents three main results. First, we measured the relative affinities of AraC for the I_1, O_2, and I_2 half-sites as well as for non-specific DNA in the presence and absence of arabinose. Second, we demonstrated that AraC possesses an intrinsic preference to loop between distally located half-sites rather than bind adjacently located half-sites. Third, we developed a thermodynamically based method using the relative affinities of AraC for the various half-sites in the calculation of the expression levels of the *ara* p_{BAD} promoter. This calculation method closely predicts the experimentally observed behavior of the promoter under a wide variety of conditions as well as predicts that AraC possesses an intrinsic preference to loop in the absence of arabinose.

Because we could describe quantitatively the regulatory behavior of the *ara* regulatory system, it seems unlikely that we are missing any of its essential regulatory components or features. The simplest description of the *araBAD* regulatory system accommodates not only known biochemical data but also our new data on the effects of various alterations to the native regulatory sequence.

Our calculation of the energetics of the various states of the *ara* system utilized the relative affinities of AraC for the various half-sites. Although measurement of relative affinities is much easier and more precise than measurement of absolute affinities, as described in the Appendix, the utilization of relative affinities required special handling. We measured the relative affinities using linear DNA, but *in vivo* the DNA is supercoiled. We expect, however, the relative effects due to supercoiling to be equal for each half-site. It must further be noted that salt conditions *in vivo* are likely to be different from those used in these assays and so the relative affinities *in vivo* could deviate from our predictions. It has been found, however, that although the absolute affinities of AraC for its DNA-binding sites vary dramatically with the salt concentration (Hendrickson & Schleif, 1984; Carra & Schleif, 1993), the relative affinities do not (Martin & Schleif, 1987).

AraC protein normally loops between the distally located O_2 and I_1 half-sites. As shown in Figure 1, these closely related asymmetric half-sites are in a direct repeat orientation on linear DNA. It seems likely, however, that in supercoiled or supertwisted DNA these sites can be located close to one another in essentially an inverted repeat orientation without large energy cost. If each DNA-binding domain of AraC were held in a fixed position with respect to its dimerization domain, then the DNA-binding domains would be positioned with the same inverted repeat symmetry as is possessed by the two dimerization domains (Soisson *et al.*, 1997). They would thus be readily available for looping between distally located direct repeat half-sites.

The preference to loop is not easily accommodated by the alternative dimerization interfaces mechanism that has been proposed based on the crystal structures of the dimerization domain of AraC that were determined in the presence and absence of arabinose (Soisson *et al.*, 1997). In a mechanism consistent with the two structures, the addition of arabinose was postulated to change the dimerization interface of the protein to one in which the connection points to the DNA binding domains are 20 Å closer to one another. Such a model explains neither the intrinsic preference of apo-AraC to engage in looping interactions nor the preference of AraC for a specific orientation between the two DNA half-sites to which it binds to loop. As described more fully in the accompanying paper (Saviola *et al.*, 1998), we now favor a mechanism for AraC action in which the N-terminal arm of AraC binds to the DNA binding domain of the protein in the absence of arabinose, and is pulled away from the domain in the presence of arabinose, thereby providing more freedom to the DNA binding domains.

Materials and Methods

Strains, media and β-galactosidase assays

All β-galactosidase and *in vivo* foot printing assays were performed using isogenic strains SH321 ($AraC^-$) or SH322 ($AraC^+$) (Hahn *et al.*, 1984). Cell cultures for β-galactosidase assays were grown in M10 medium (Schleif & Wensink, 1981) plus or minus arabinose 2% (w/v) at 37°C with shaking to an A_{550} between 0.3 and 0.9. β-Galactosidase expression was measured by the method of Miller as described (Maniatis *et al.*, 1982). Cell cultures for *in vivo* foot printing were grown in M10 medium plus or minus arabinose 2% (w/v) at 37°C with shaking to an A_{550} between 0.3 and 0.5.

Synthesis of operator DNA

The wild-type operator and artificial operator constructs were amplified by polymerase chain reaction (PCR), from plasmid DNA template. I_1-I_2 (sequence, half-sites underlined: TAGCATTTTTATCCATAAGATTAGCGATCCTACCTGA) was derived from pES51 (Huo *et al.*, 1988) using oligos of sequence AAGTCCACATCGAT-

TATTTGC and GTGTTCACCAATCAAATTCAC, I_1-I_1; (sequence, half-sites underlined: TAGCATTTTTATCCATAAGATTAGCATTTTTATCCATA) from pP3-10-I_1I_1 (Reeder & Schleif, 1993) using oligos of sequence CAAATAGGGGTTCCGCGCAC and GTGCGCGTGCAGCCCTTATTGCCCGG; and I_1-*ns* (sequence, half-sites underlined: TAGCATTTTTATCCATAGATCCTGGTACCGAATTCAG) from pP3-ΔI_D (Reeder & Schleif, 1993) using oligos of sequence CAAATAGGGGTTCCGCGCAC and GTATCTAGAAACAGTAGAGAG. The I_1-O_2 template plasmid pI_1-O_2 (I_1-O_2 sequence, half-sites underlined, TAGCATTTTTATCCATATCTAGAAACCAATTGTCCATA) was constructed as a series of four overlapping oligonucleotides that were fused using PCR ligation, cut with *Eco*RI, then inserted into the *Eco*RI site of pUC19. The complete sequence of the insert is GAATTCGATCAGACATTGTCTAGACGATCAGACATTGTGCACATCGATACGTAGTACGCGTAAAAACGCGCAAAAACCGCGGTCATATAGCATTTTTATCCATAAGAAGAAACCAATTGTCCATAAGATCTCAGACAGTAGAGTCGACACGATCAGACATTGGATCCTCAGACATGAGCTCGCATCATGAATTC. I_1-O_2 was then extracted with PCR using oligos of sequence CATCAGGATTCGATCAG and GTAGTCGAATTCATG. For the DNA migration retardation assays the operator synthesis PCR was performed with one primer ^{32}P-labeled and the other primer in excess to minimize the amount of labeled single-stranded product. For the DNA migration retardation assay it is necessary that bound and free bands of all DNA species migrate at different rates such that each is separated from the others. We achieved proper separation using fragment lengths for I_1-I_2 of 440 bp, I_1-O_2 of 206 bp, I_1-I_1 of 268 bp, and I_1-*ns* of 255 bp, and operators positioned 30 bp from the fragment end for I_1-I_2 and centered on the latter three fragments.

DNA migration retardation assay

The DNA migration retardation assay, in which AraC must bind to two half-sites to remain on the DNA for the duration of the electrophoresis, was performed essentially as described (Hendrickson & Schleif, 1984) with the exception that two species of labeled DNA fragments, of different lengths and predetermined not to co-migrate in either bound or free forms, were added to the assay mix for a given sample. AraC protein was purified to homogeneity by Jeff Withey (Schleif & Favreau, 1982). All assay binding reactions were performed in 20 µl volumes. The buffer used for the binding reaction contained 150 mM KCl, 10 mM Tris-acetate (pH 7.4), 1 mM EDTA, 5% glycerol, 1 mM dithioerythritol, and 0.05% NP 40 (v/v). Samples containing arabinose did so at 1% (w/v). AraC protein was diluted immediately prior to use by slowly adding binding buffer to an aliquot of protein stock solution. AraC was added to a final concentration between 1 and 100 nM. DNA fragments were added to concentrations of 1 to 10 nM. Samples were incubated for 20 minutes at 37°C, a time predetermined to ensure the reactions had reached equilibrium. Electrophoresis through rinsed, presoaked 6% (w/v) acrylamide, 0.1% (w/v) bis-acrylamide gels containing 10 mM Tris-acetate, 1 mM K-EDTA was carried out at 5 V/cm for one to three hours in a horizontal apparatus with a connected recirculating pump maintaining the buffer temperature at 20°C. Gels were vacuum dried and the radioactivity of bands was quantified using a Molecular Dynamics Phosphorimager PC™.

Numerical solution of descriptive equations

Equations describing the behavior of the *ara* regulatory region were solved numerically using the Newton method as implemented in Microsoft Excel®.

Upstream regulatory sequence mutants

Operator mutant-*lacZ* fusion reporter plasmids were constructed by standard molecular biological techniques (as described by Maniatis *et al*., 1982). The *galK* reporter gene of pES51 was replaced with *lacZ* gene DNA amplified from plasmid pP7 (Reeder & Schleif, 1993). O_2 was removed from pES51 by restriction cleavage at the *Sal*I and *Nco*I sites and replaced with half-sites made from overlapping oligonucleotides designed so as to hybridize leaving sticky ends with complementary sequence to the overhangs left at those restriction sites. The sequences inserted were (with the half-site sequence underlined): for I_1 CATGGAGAGAATAGCATTTTTATCCATATTGCATCAGACATTG, for I_2 CATGGAGAGAATAGCGGATCCTACCTGATTGCATCAGACATTG. The O_2 deletion was created by filling out the cut ends of the double cut pES51 above with Klenow fragment and blunt-end ligation.

The I_1-O_2 construct used for *in vivo* footprinting was created by inserting the pES51 *Eco*RI to *Hin*dIII fragment, containing the wild-type upstream regulatory sequence, into pUC19 and then altering the normal I_2. This alteration was performed by replacing the *Nhe*I to *Hin*dIII fragment from the pUC19-I_1-I_2 with an overlapping oligonucleotide pair as above with the sequence for I_1-O_2: CTAGCCCATAGCATTTTTATCCATATCTAGAAACCAATTGTCCATA. The reversed O_2 construct is identical to the I_1-O_2 construct above with the exception that the *Nco*I to *Sal*I fragment reads CATGGAGAGATATGGACAATTGGTTTCTTTGCATCAGACATTGT.

In vivo footprinting

To a 10 ml cell culture was added 10 µl dimethyl sulfate (DMS) with continued shaking for two minutes. The culture mix was spun down and the DNA was harvested by mini-prep. All of the resulting DNA was reconstituted in 70 µl distilled water. The DNA was cleaved with 10 µl piperidine for 30 minutes at 90°C then extracted with butanol and rinsed with 95% ethanol. The extracted DNA was resuspended in 20 µl distilled water, 5 µl aliquots of which were used as templates for extension reactions with 5 ng each of ^{32}P-labeled primers complementary to the bottom strand. The primer sequences were: for I_1-O_2 AAAAGTGTCTATAATCACGGC; for distal O_2 CGTTGTAAAACGACGGCCAGT; and for reversed O_2 AATAAGCGGGGTTACCGGTTGGG. The products of the extension reactions were run out on a 6% denaturing gel of the type used for sequencing.

Acknowledgments

We gratefully acknowledge John Carra for his preliminary work in this study. This work was supported by National Institutes of Health grant GM18277 to R.F.S.

References

Ackers, G. K., Johnson, A. D. & Shea, M. A. (1982). Quantitative model for gene regulation by lambda phage repressor. *Proc. Natl Acad. Sci. USA,* **79**, 1129–1133.

Brenowitz, M., Jamison, E., Majumdar, A. & Adhya, S. (1990). Interaction of the *Escherichia coli* Gal repressor protein with its DNA operators *in vitro*. *Biochemistry,* **29**, 3374–3383.

Brenowitz, M., Pickar, A. & Jamison, E. (1991). Stability of Lac repressor mediated looped complex. *Biochemistry,* **30**, 5986–5998.

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

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

Dalma-Weiszhausz, D. D. & Brenowitz, M. (1992). Interactions between DNA-bound transcriptional regulators of the *Escherichia coli gal* operon. *Biochemistry,* **31**, 6980–6989.

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.

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.

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

Hahn, S., Dunn, T. & Schleif, R. F. (1984). Upstream repression and CRP stimulation of the *Escherichia coli* L-arabinose operon. *Proc. Natl Acad. Sci. USA,* **81**, 5017–5020.

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). Alternative DNA loops regulate the arabinose operon in *Escherichia coli*. *Proc. Natl Acad. Sci. USA,* **85**, 5444–5448.

Lee, D. & Schleif, R. F. (1989). *In vivo* DNA loops in *araCBAD:* size limits and helical repeat. *Proc. Natl Acad. Sci. USA,* **86**, 476–480.

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.

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

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

Martin, K., Huo, 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 in these same sites. *Proc. Natl Acad. Sci. USA,* **83**, 3654–3658.

Martin, K. J. & Schleif, R. F. (1987). Equilibrium DNA-binding of AraC protein. *J. Mol. Biol.* **195**, 741–744.

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**, 205–213.

Saviola, B., Seabold, R. & Schleif, R. (1998). Arm-domain interactions in AraC. *J. Mol. Biol.* **278**, 539–548.

Schleif, R. F. (1992). Regulation of the L-arabinose catabolic operon *araBAD*. In *Transcriptional Regulation* (McKnight, S. & Yamamoto, K., eds), pp. 643–665, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schleif, R. F. & Favreau, M. A. (1982). Hyperproduction of AraC protein from *Escherichia coli*. *Biochemistry,* **21**, 778–782.

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

Senear, D. F. & Brenowitz, M. (1991). Determination of binding constants for cooperative site-specific protein-DNA interactions using the gel mobility-shift assay. *J. Biol. Chem.* **266**, 13661–13671.

Soisson, S. M., MacDougal-Shackleton, B., Schleif, R. F. & Wolberger, C. (1997). Structural basis for ligand-regulated oligomerization of AraC. *Science,* **276**, 421–425.

Zhang, X., Reeder, T. & Schleif, R. F. (1996). Transcription activation at *ara* p_{BAD}. *J. Mol. Biol.* **258**, 14–24.

Appendix

Calculation of *araBAD* regulation

In the same way that AraC will partition itself between two DNA molecules in accordance with the strength of its binding to the two, AraC will partition itself amongst the sites constituting the regulatory region of *ara* p_{BAD} in accordance to its effective binding affinities. Preliminary calculations showed that the most highly occupied states would be the following: free DNA, DNA with AraC bound at I_1 and O_2, denoted I_1O_2, DNA with AraC bound at I_1 and I_2, denoted by I_1I_2 and DNA with AraC bound at I_1 and at some non-specific DNA position, denoted I_1ns (Figure 5). We assume that the system is in steady state and that the activity of the *ara* p_{BAD} promoter is proportional to the amount of the DNA in the I_1I_2 state.

Let:

K_{I1} be the affinity of AraC for the I_1 half-site relative to its affinity for non-specific DNA in the absence of arabinose,

K_{I1arab} be the affinity of AraC for the I_1 half-site relative to its affinity for non-specific DNA in the presence of arabinose,

K_{O2} be the affinity of AraC for the O_2 half-site relative to its affinity for non-specific DNA in the absence of arabinose,

K_{O2arab} be the affinity of AraC for the O_2 half-site relative to its affinity for non-specific DNA in the presence of arabinose,

K_{I2} be the affinity of AraC for the I_2 half-site relative to its affinity for non-specific DNA in the absence of arabinose,

K_{I2arab} be the affinity of AraC for the I_2 half-site relative to its affinity for non-specific DNA in the presence of arabinose,

K_{ns} be the affinity of AraC for the *ns*, nonspecific, half-site relative its affinity for non-specific DNA, which equals 1 by definition in the presence or absence of arabinose,

N be the number of non-specific half-sites to which AraC can loop to from I_1,

$$\frac{D_{I1I2}}{D + D_{I1I2} + D_{I1O2} + D_{I1-ns}} = \frac{P \times K_{I1} \times K_{I2} \times T \times C_{-a}}{1 + P \times K_{I1} \times K_{I2} \times T \times C_{-a} + P \times K_{I1} \times K_{O2} \times T \times L_{-a} + N \times P \times K_{I1} \times K_{ns} \times T \times L_{-a}}$$

In the presence of arabinose, the fraction in this I_1I_2 state is:

$$\frac{aD_{I112}}{aD + aD_{I112} + aD_{I102} + aD_{I1-ns}} = \frac{P \times K_{I1arab} \times K_{I2arab} \times T \times C_{+a}}{1 + P \times K_{I1arab} \times K_{I2arab} \times T \times C_{+a} + P \times K_{I1arab} \times K_{O2arab} \times T \times L_{+a} + N \times P \times K_{I1arab} \times K_{ns} \times T \times L_{+a}}$$

T, be the conversion from the half-site relative affinities to actual affinity of AraC for two adjacent half-sites (thus the binding affinity for AraC to I_1-I_2 in the absence of arabinose equals $K_{I1} \times K_{I2} \times T$),

L_{-a} be a factor to include any overall preference for or against forming a loop in the absence of arabinose (this factor includes the intrinsic preference of AraC to loop and the possible consequences of interactions with RNA polymerase and CAP; thus, the overall affinity for looping between I_1 and O_2 in the absence of arabinose is $K_{I1} \times K_{O2} \times T \times L_{-a}$),

L_{+a} be like L_{-a} but in the presence of arabinose,

C_{-a} be whatever additional propensity or cost is involved in AraC binding to two adjacent half-sites minus arabinose (from the definition of T, $C_{-a} = 1$),

C_{+a} be like C_{-a} but in the presence of arabinose,

P be the effective concentration of free AraC protein per cell (one molecule free per cell would be 10^{-9} M),

D be the concentration of *ara* regulatory region DNA free of bound AraC,

D_{xy} be the concentration of *ara* regulatory region DNA with AraC bound at the half-sites x and y and,

aD_{xy} be like D_{xy} but in the presence of arabinose.

For a binding reaction of $A + B$ producing AB described by an affinity K_{AB}:

$$K_{AB} = \frac{[AB]}{[A] \times [B]}$$

$$[AB] = [A] \times [B] \times K_{AB}$$

Analogously, for the AraC binding half-sites in the absence of arabinose:

$$D_{I1O2} = [P] \times [D] \times K_{I1} \times K_{O2} \times T \times L_{-a'}$$

$$D_{I1I2} = [P] \times [D] \times K_{I1} \times K_{I2} \times T \times C_{-a'}$$

and:

$$D_{I1-ns} = [P] \times [D] \times K_{I1} \times K_{ns} \times T \times L_{-a}$$

The fraction of the operons in a culture that are in the I_1I_2 state, for example, equals:

Equations similarly may be written for mutant situations. If O_2 is deleted, then the D_{I1O2} state is absent. If O_2 is changed to I_1, then the term K_{O2} is changed to K_{I1}, etc. The only states involving non-specific binding appear to involve binding at I_1 and a nonspecific site. We estimate that only about ten such sites can be involved in such binding and hence $N = 10$.

From the data presented in Table 1, the following values may be derived:

$$K_{I1} = 680,\ K_{O2} = 72,\ K_{I2} = 4,\ K_{ns} = 1$$

and:

$$K_{I1arab} = 700,\ K_{O2arab} = 200,\ K_{I2arab} = 10,\ K_{ns} = 1$$

By definition $C_{-a} = 1$. From measurements of AraC binding, the presence of arabinose increases the affinity about 40-fold, thus $C_a = 40$. P is of the order of 10^{-9} to 10^{-10}.

This leaves T, L_{-a}, and L_{+a} to be determined.

Using the facts that arabinose induces the wild-type system approximately 100 × , that this induction is about half of the maximum possible, and that in an O_2 deletion strain the basal level is elevated about fourfold, we can generate the following requirements:

in the presence of arabinose:

$$\frac{aD_{I112}}{aD_{total}} = 0.5$$

in the absence of arabinose:

$$\frac{D_{I112}}{D_{total}} = 0.005$$

and in a ΔO_2 strain in the absence of arabinose:

$$\frac{\dfrac{D_{I112}}{D + D_{I112} + D_{I1-ns}}}{\dfrac{D_{I112}}{D + D_{I1O2} + D_{I112} + D_{I1-ns}}} = 4$$

The complete set of values used in our computations are as follows:

$K_{I1} = 680,\ K_{O2} = 72\ K_{I2} = 4,\ K_{ns} = 1$

$K_{I1arab} = 700,\ K_{O2arab} = 200\ K_{I2arab} = 10,\ K_{ns} = 1$

$N = 10$

$C_{-a} = 1$

$C_{+a} = 40$

$P = 10^{-10}$

$T,\ L_{-a'}$ and L_{+a} were solved for.

Numerically solving this system of three equations yields:

$$T = 130,000,\ L_{-a} = 8.3,\ \text{and}\ L_{+a} = 1.3$$

It is noteworthy that the product of T, K_{I1} and, K_{I2} yields a value for the true K_a of AraC binding to I_1-I_2 that is consistent with the best data available from direct measurement (Hendrickson & Schleif, 1984; Carra & Schleif, 1993).

References

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

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.

Edited by R. Ebright

(Received 26 August 1997; received in revised form 10 February 1998; accepted 10 February 1998)