Alternative DNA loops regulate the arabinose operon in *Escherichia coli* (repression/AraC protein/DNA–protein interaction/loop equilibria/regulation model)

**Li Huo, Katherine J. Martin*, and Robert Schleif**

Department of Biochemistry, Brandeis University, Waltham, MA 02254

Communicated by Peter H. von Hippel, April 4, 1988 (received for review December 17, 1987)

**ABSTRACT** The *araCBAD* regulatory region of *Escherichia coli* contains two divergently oriented promoters and three sites to which AraC, the regulatory protein of the operon, can bind. This paper presents the results of in vivo dimethyl sulfate “footprinting” experiments to monitor occupancy of the three AraC sites and measurements of activity of the two promoters. These measurements were made both in the absence of the inducer arabinose and at various times after arabinose addition to growing cells containing the wild-type ara regulatory region or the regulatory region containing various deletions and point mutations. The data lead to the conclusion that two different DNA loops can form in the *ara* regulatory region. These loops are generated by AraC protein molecules binding to two different DNA sites and binding to each other. One of these loops predominates in the absence of arabinose and plays a major role in repressing activity of one of the promoters. Upon the addition of arabinose the amount of the first loop type, the repression loop, decreases and the amount of a second loop increases. Formation of this second loop precludes the counterproductive formation of the repression loop.

AraC protein is a multifunctional regulatory protein of the genes required for uptake and catabolism of L-arabinose in *Escherichia coli* (1–4). This paper concerns the inducing and repressing activities of AraC on the two promoters located in the regulatory region between the *araC* and *araBAD* genes. One, *P*$_{BAD}^*$, serves the *araBAD* catabolic genes, and the other, *Pc*, serves the *araC* regulatory gene. In the presence of arabinose, AraC protein bound at the *araI* site, which is immediately adjacent to the RNA polymerase binding site of the *P*$_{BAD}^*$ promoter, stimulates transcription of the *araBAD* genes (Fig. 1). The protein also represses mRNA synthesis from *P*$_{BAD}^*$ in the absence of arabinose by a mechanism requiring the formation of a DNA loop (5–7). One end of this loop requires AraC protein binding at the *araO2* site, an operator that is located >200 base pairs (bp) away from the start of *araBAD* transcription. Genetic experiments suggest that the other end of the loop is AraC protein bound to the *araI* site (6). AraC protein represses its own synthesis from the promoter *Pc*, most likely by binding to a third AraC protein binding site in the regulatory region, the operator *araO1*, and directly blocking RNA polymerase entry to the promoter for synthesis of AraC protein, *Pc* (8, 9).

A variety of experiments have led to the conclusion that looping occurs in the *ara* system (5–7, 10). Additionally, DNA looping appears to be a common mechanism for gene regulation, as a number of other systems also show one or more of the properties displayed by the *ara* system and can be interpreted to loop so that two proteins bound to separated sites on DNA are in direct contact (11–22). Finally, the

![Fig. 1. Schematic drawing of the regulatory region of the *araCBAD* operon, showing the *araC* binding sites, *araI*, *araO1*, and *araO2*, as well as portions of the flanking *araC* and *araBAD* genes. Numbering of base pairs is relative to the *P*$_{BAD}^*$ transcription start site at +1. Solid, AraC protein; open, cAMP receptor protein (CRP); cross-hatch, RNA polymerase.](image)

strength of the protein–DNA interactions and the protein–protein interactions in some systems (14, 16, 20, 22) permits loops generated on linear DNA to be observed in the electron microscope.

Previous in vivo dimethyl sulfate methylation “footprinting” studies of the *araCBAD* operon revealed that in the absence of the two other AraC protein binding sites, *araO2* binds AraC protein so weakly that it is virtually unoccupied (9). That is, the presence of the other sites, even though they are located >100 bp away, greatly increases AraC protein occupancy of *araO2*. Since AraC bound at *araO2* must be interacting with one or both of the other sites to generate this binding-site cooperativity, either they are communicating by sending a signal along the DNA between the sites or, more likely, the DNA is looped to permit the proteins bound to each site to interact directly.

Is *araI* the sole source of the binding-site cooperativity leading to AraC occupancy of *araO2*? Here we report that AraC bound to *araO2* interacts not only with AraC bound to *araI* but also with AraC bound to *araO1* and, therefore, participates in the formation of an alternative loop in the *araCBAD* regulatory region. We find that both types of loops are involved in regulating both the *Pc* and *P*$_{BAD}^*$ promoters. We also present data suggesting that only one or the other of the loops can exist at one time, that arabinose shifts the amounts of the two looped forms present, and that looping may restrict access of AraC protein and RNA polymerase to binding sites within the looped region.

Abbreviations: CRP, cAMP receptor protein; *P*$_{BAD}^*$, promoter for *araBAD* operon; *Pc*, promoter for *araC* gene.

*Present address: Department of Biochemistry and Molecular Biology, 7 Divinity Avenue, Harvard University, Cambridge, MA 02138.*

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

5444
MATERIALS AND METHODS

Media, Strains, and General Methods. The isogenic AraC+ and AraC− strains were SH322 and SH321 (6). Activity of P_BAD was monitored by measuring galactokinase (GalK) encoded by a plasmid in which wild-type or mutant araCBAD regulatory region was fused to galK (5). Cells containing the plasmid were grown in M10 minimal medium (23) at 35°C and assayed as described (6). Araabinose, when added, was at 0.2%. The activities of P_BAD and P_C were also measured by fusing the araCBAD regulatory region in either orientation to lacZ and quantitating β-galactosidase. The araO1 deletion mutant KM270 was constructed by deleting 27 bp in the araO1 region, positions −119 to −146 with respect to the transcription start of P_BAD at +1. Six base pairs of filled-out EcoRI linker have been inserted at the deletion site and 21 bp of compensating DNA have been inserted upstream at the BstEII site. Thus, the wild-type spacing between araL and araO1 has been maintained. The previous sequence at the BstEII site was GTGTTAACC, and the new sequence with the 21-bp insertion is GTGAACTCTAGATCGATCTAGGTAACC. When β-galactosidase activity was to be measured, the cells were grown at 30°C in M10 minimal medium containing 0.2% casein amino acids and 0.2% glycerol. Cells were assayed (24) with the following modifications. Samples were treated with 20 μl of 0.1% NaDodSO4 and 40 μl of chloroform instead of toluene and enzyme incubations were at 32°C instead of 28°C.

In Vivo Footprinting. Methods were as described (7). All the footprint experiments utilized the plasmid pTD3 (5) or derivatives, in which the 440-bp complete araCBAD regulatory region is fused to the galK gene. One of the derivatives contained the wild-type-like ara regulatory DNA, ES51, in which restriction endonuclease cleavage sites have been created between regulatory sites to facilitate the manipulation of the DNA. ES51 differs from the wild-type ara regulatory region at the following positions: G→C change at −76, T→G at −77, T→C at −107, C→G at −162, A→C at −166, T→G at −219, C→A at −220, T→A at −248, C→T at −251, C→G at −290, and T→G at −291. These changes lie outside known protein binding sites and do not significantly alter regulation of P_C or P_BAD. The deletions used in the in vivo footprinting experiments were made by cutting ES51 with HindIII and another restriction enzyme. The phage M13 replication origin was inserted into pTD3 and a derivative containing the araL-site point mutation BR333 (25). Gel binding assays (9) have shown that this mutation reduces the affinity of araL for AraC protein by a factor of 30 by increasing the dissociation rate. The M13 origin facilitated DNA sequencing and the oligonucleotide site-directed mutagenesis that was used to construct the araO3 point mutation LH7 and the araL–araO3 double point mutant plasmid pLH6. The occupancy of mutated araO3, LH7, was reduced to about one-third of normal as measured by in vivo footprinting (data not shown).

P_C Expression Kinetics. All plasmids were derivatives of plasmid pTD4 (5), in which transcription of galK is driven by P_C. For the assay of the time course of P_C activity, the ara region of pTD4 was fused to the lacZ gene of pDL3 and cells were grown at 35°C for at least four generations. Aliquots of 10 μl of cells were taken every 2 min for β-galactosidase assays, and 1-ml aliquots were taken every 12 min for measurement of cell density.

RESULTS

Deletions Identify araO1 as Aiding araO2 Occupancy. To localize the source(s) of the cooperativity that assists binding of AraC protein to araO2, we removed progressively larger amounts of the ara regulatory region from the araL end (Fig. 1). The results showed that either araO1 or araL could provide the interactions necessary for AraC protein to occupy araO2; that is, looping could occur from araO2 to araO1 or from araO2 to araL.

First, we shall discuss the results obtained in the absence of arabinose. Previous in vitro and in vivo experiments showed that the increase in the methylation rates of a guanine by dimethyl sulfate in AraC binding sites is a sensitive indication of the occupancy of the sites by AraC protein (5, 7). Although the precise relationship between the occupancy and increase in methylation rate is not known, it seems likely to be linear. Deletions of the araL or araL–CRP sites had only small effects on the apparent occupancy of araO2 (Fig. 2A). This shows that neither the araL nor the CRP binding site is solely required for the cooperativity that gives AraC binding at araO2 (looping). A still larger deletion, one removing araO1 in addition to the CRP site and araL, almost completely eliminated binding to araO2. This result shows that the looping present in the araL and araL–CRP site deletion strains involves araO1.

In the presence of arabinose, the apparent occupancy of araO2 was reduced by a factor of 2–3 on the wild-type plasmid. Upon deletion of araL, occupancy of araO2 was further reduced. The occupancy of araO2 was still further reduced upon deletion of araL and the CRP binding site and was eliminated when araL, the CRP site, and araO1 were all

![Fig. 2. In vivo dimethyl sulfate footprinting (7) of AraC (C PROT) binding to the araO2 region in deletion and point mutation plasmids in the presence or absence of arabinose as indicated. (A) Below the autoradiogram, the ara regulatory region in the various deletion plasmids is shown to scale. Numbers at left indicate nucleotide positions relative to the araBAD transcription start site. WT, wild-type-like ara regulatory DNA ES51; (B) WT, wild-type ara regulatory DNA; O1, DNA with the araO1− point mutation (LH7), a G→T change at position −110; O1−T, LH6 DNA, which has the araO1− mutation combined with the araL− mutation BR333 [an A→T change at position −57 (25)].]
deleted. These results provide evidence for the existence of both the araO2-araO1 loop and the araO2-araO1 loop.

Point Mutations Identify aral as Aiding araO2 Occupancy. The experiments described above clearly identified araO1 as a source of the interactions that enable araO2 to be occupied by AraC protein. Those experiments less clearly identified the expected site, aral, also as a source of these interactions. To strengthen the evidence for looping between araO2 and aral, we mutated araO1 to reduce AraC binding there and then compared the occupancy of araO2 in araO1-araO2 strains and araO1-araO1 strains. The difference in araO2 occupancy between these two strains will then correspond to the binding-site cooperativity contributed by only aral.

The strain with the mutant araO1 site, LH7, showed no reduction in looping as inferred from araO2 occupancy (Fig. 2B). Presumably aral provided the cooperative interactions necessary for araO2 occupancy. When a mutant aral, BRS33, was introduced into the strain, we observed significant reduction in araO2 occupancy. Thus, this set of experiments showed that looping can also occur between araO2 and aral.

The results shown in Fig. 2 indicate that, in appropriate deletion or point mutation strains, looping occurs from araO2 to either araO1 or aral in the presence or absence of arabinose. The decrease by a factor of 2–3 in the extent of the altered methylation pattern at araO2 in the presence of arabinose likely results from reduction of looping and is discussed later. By themselves, these results do not indicate which loop might normally predominate in the presence or absence of arabinose, but they show that when formation of one loop is prevented, the remaining loop can form in the presence or absence of arabinose.

araO1 Indirectly Affects Regulation of P_{BAD}. We examined P_{BAD} activity after damaging araO1. This site is upstream from all the sequences required for activity of P_{BAD} in the absence of the araO2 site. That is, deletions entering from upstream do not reduce transcription from P_{BAD} until they reach about 20 bp past araO1 (26). However, reducing AraC protein binding to araO1 while retaining the presence of araO2 reduced the activity of P_{BAD}. We damaged araO1 in two ways: (i) by introducing the double mutation previously mentioned that reduces the apparent in vivo occupancy of araO1 by a factor of 3 and (ii) by deleting araO1. The resulting activities of P_{BAD} were decreased (Table 1).

Under both noninducing and inducing conditions the effect of the araO1 deletion was greatest when araO2 and aral were correctly oriented with respect to each other around the DNA helix (data not shown).

aral Indirectly Affects Regulation of P_C. It is reasonable that AraC protein should regulate its own synthesis by an autorepression scheme, and a plausible mechanism for achieving this is available since araO2 overlaps P_C. Less obvious either in a mechanistic or in a physiological role is the fact that for about 10 min after the addition of arabinose to cells, P_C activity is about 5-fold higher than normal (27, 28). After this transient period of hyperactivity, the promoter activity falls to within a factor of 2 of the value it had before the addition of arabinose.

To test whether P_C, like P_{BAD}, might be affected by the ara loops, we altered looping in the ara regulatory system without directly affecting the P_C promoter or sequences transcribed from it. This was done by deleting aral so that only the araO2-araO1 loop should be able to form. Not only was the transient hyperactivity eliminated, it was reversed and the activity of P_C decreased after arabinose addition (Fig. 3).

To explore more carefully the transient hyperactivity of P_C, we measured apparent araO1 and araO2 occupancy during the period following arabinose addition. Fig. 4 shows that araO1 was poorly occupied before addition of arabinose and that after arabinose addition its occupancy increased over a period of about 10 min. After arabinose addition, apparent araO2 occupancy dropped quickly to about one-quarter of its pre-arabinose value, and then over about 10 min araO2 apparent occupancy rose to about half its pre-arabinose value.

DISCUSSION

Summary and Model. Four basic experimental results are presented in this paper. (i) Both the aral site and the araO1 site potentiate binding by AraC protein at the much weaker AraC protein binding site, araO2. (ii) Damaging araO1, a site not in P_{BAD}, lowers the uninduced and induced levels of P_{BAD} activity. (iii) Damaging aral, which lies well upstream of P_C, eliminates the 10-min period of P_C hyperactivity that normally follows arabinose addition. (iv) After arabinose addition, apparent araO1 occupancy by AraC increases over a 10-min interval during which apparent occupancy of araO2 rapidly decreases and then increases again.

In light of the complexity of the ara system, we have found it most efficient first to present a model for ara regulation that is suggested by the data presented here and previously. Although this model undoubtedly is an oversimplification and is not fully proven, it serves as a convenient summary of much experimental data.

In the absence of arabinose, most copies of the ara regulatory region contain a loop between araO2 and aral mediated by AraC protein bound to both of these sites (Fig. 5). This loop prevents AraC protein bound to aral from

![Fig. 3. Activity of P_C in P_{BAD}-lacZ fusion strains after addition of arabinose to aral- (wild-type) and aral-deleted cells. The plasmids were ESS1 and ESS1 deleted from the BamH1 site (located at -47) to the NheI site (located at -81).](image-url)
This loop or AraC occupancy of either araO₁ or araO₂ represses $P_C$.

**Discussion of the Data.** We have used *in vivo* footprinting techniques to show that the binding of the AraC regulatory protein to its upstream operator site, $araO_2$, requires cooperative interactions that can be provided by either of two other AraC binding sites, $araO_1$ or $araI$, implying the existence of two aral loops.

One of the $ara$ loops, that involving AraC bound at $araO_2$ and $araI$, was previously shown to generate repression of *araBAD* (5–7). This repression occurs in the absence of the inducer arabinose and reduces the level of $araBAD$ transcript by a factor of about 8. Apparently, this loop prevents or substantially inhibits AraC bound at the induction site $araI$ from entering its inducing state.

The second $ara$ loop, that involving AraC bound at $araO_2$ and $araO_1$, was previously unknown. The existence of this second loop suggests a regulatory scheme in which one loop generates $araBAD$ repression and the other loop prevents formation of the first loop, thereby assisting induction. Therefore, we tested whether removing the $araO_1$ site affects $araBAD$ expression. In support of the idea, our results show that damaging $araO_1$ reduces both the induced and the uninduced expression of $P_{BAD}$, suggesting that the change increases the amount of the repression loop under noninducing, inducing conditions.

The addition of arabinose to cells initiates a 10-min period of hyperactivity of the araC promoter (27). This previously unexplained phenomenon can be understood in light of the alternative-looping scheme and the footprinting data presented here. The rapid loss of $araO_2$ occupancy upon addition of arabinose indicates that the repression loop opens. This would generate a temporary overpopulation of the unlooped state in which $araO_1$ would be largely unoccupied and in which RNA polymerase should have free access to $P_C$. During this period, $P_C$ would be hyperactive and $araO_2$ would be unoccupied, presumably due to the absence of looping. Then, after about 10 min, when we observe $araO_1$ and $araO_2$ to become occupied, indicating formation of the $araO_2$–$araI$ loop, the hyperactivity of $P_C$ should, and does, cease. Deleting $araI$ has the predicted effect on the transient hyperactivity of $P_C$. In the absence of $araI$, there should be no transient overpopulation of the unlooped state and hence no transient hyperactivity, and indeed, we saw none.

It is of interest to estimate quantitatively the amounts of the repression loop and of the second, "antirepression" loop present during noninducing and inducing conditions. The approximate levels of the two types of loops can be calculated from the available data on the activity of $P_{BAD}$. The results of such calculations (Table 2) show that, subject to the simplifying assumptions mentioned in the legend, arabinose addition decreases the fraction of molecules containing the $araO_2$–$araI$ repression loop from 90% to 33% and increases the fraction of molecules containing the $araO_2$–$araI$ antirepression loop from 5% to about 40% (Fig. 5).

**Additional Considerations.** Two types of looping are formally possible amongst $araI$, $araO_1$, and $araO_2$. One is a double looped structure simultaneously involving several dimers of AraC protein and all three DNA sites, and the other is the mutually exclusive situation in which only one loop may exist at any one time. If a complex involving all three DNA sites formed, then eliminating one of the sites would seem likely to weaken the structure and lessen the probability that the remainder of the looped structure would form. On the other hand, with alternative looping, eliminating one loop form would increase the amount of the other loop form. Since we observed the latter for either loop, we favor models based on alternative looping as we have described above.

We might worry that forming the 140-bp $araO_2$–$araO_1$ loop would be energetically difficult. Although unstrained circles
Table 2. Inducibility and looping frequencies

<table>
<thead>
<tr>
<th>Strain</th>
<th>Relative $P_{BAD}$ activity</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$araO_2$-$araO_1$ loop</td>
<td>$araO_2$-$araO_1$ loop</td>
</tr>
<tr>
<td>Wild type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uninduced</td>
<td>1</td>
<td>0.90</td>
</tr>
<tr>
<td>Induced</td>
<td>100</td>
<td>0.33</td>
</tr>
<tr>
<td>$araO_2$ deletion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uninduced</td>
<td>10</td>
<td>–</td>
</tr>
<tr>
<td>Induced</td>
<td>150</td>
<td>–</td>
</tr>
<tr>
<td>$araO_1$ deletion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uninduced</td>
<td>0.5</td>
<td>0.95</td>
</tr>
<tr>
<td>Induced</td>
<td>70</td>
<td>0.53</td>
</tr>
</tbody>
</table>

The input data are the set of numbers of $P_{BAD}$ activity. These are the typical activities of $P_{BAD}$ reported previously (6) except for the induced activity when $araO_2$ is deleted, for which its value was found to be 150 (ref. 29 and Dong-Hee Lee and Rob Lobell, personal communication). Entries for the looped or unlooped states were calculated by building (“bootstrapping”) from the activities of the promoter with the assumptions that $P_{BAD}$ is inactive on molecules containing an $araO_2$-$araO_1$ loop and that the partitioning of the molecules among the various states is governed by equilibrium constants that are affected by the presence of arabinose. For example, the amount of the $araO_2$-$araO_1$ loop in uninduced wild-type cells was calculated to be 0.9 from the fact that deletion of $araO_2$ increases the uninduced level by a factor of 10. The estimate that the remaining 0.1 of the molecules are evenly split between the $araO_2$-$araO_1$ loop and the unlooped state derives from the factor of 2 reduction in $P_{BAD}$ activity upon deletion of $araO_1$. Two decimal places of precision are presented to assist the reader in reproducing the calculations, not because we believe such precision is experimentally significant.

as small as 120 bp can be formed with DNA containing bent sequences (30), the $ara$ appears not to contain such sequences. Of course, if the protein were flexible, then a significant portion of the 360° bending around a loop could be provided by the protein, and the DNA would need to bend less and less as the loop size decreased. Similar flexibility in the protein may help explain the small (52-bp) loop size that has been observed with $\lambda$ phage repressor (13).

The activity of $P_C$ is repressed both before the addition of arabinose and more than 10 min after the addition of arabinose. This repression could come from three sources: (i) direct competition between binding of AraC to $araO_2$ and binding RNA polymerase to $P_C$, since $araO_1$ and $P_C$ overlap; (ii) RNA polymerase binding to $P_C$ being hindered by the $araO_2$-$araO_1$ loop; and (iii) AraC bound at $araO_2$ blocking elongation by RNA polymerase. The $araO_2$ site overlaps the RNA polymerase binding site for $P_C$, and one set of experiments showed that binding to these two sites was mutually exclusive (9). RNA polymerase binding at $P_C$, which lies in the middle of the $araO_2$-$araO_1$ 210-bp loop, should be sterically hindered by the existence of the loop. In an analogous situation DNase could not easily cleave phosphodiester bonds on the inside face of DNA circles of dimensions similar to those of the loop (31).

Finally, we note that the addition of arabinose shifts the loops from $araO_1$ toward $araO_2$. That is, AraC protein bound to these two sites behaves differently. Although having the DNA sequence to which a protein is bound act as an allosteric effector of the protein is not a familiar concept in biochemistry, AraC protein does in fact change its ability to repress $P_{BAD}$ when bound to a mutant $araO_1$ site (7).

We thank Sharon Jenkins for constructing E515; Dong Hee Lee for his gift of PDL3; and Pieter Wensink, Alan Brunelle, Jim Tobin, John Carra, Robert Lobell, Dong Hee Lee, and Ranjan Sen for discussions and comments on the manuscript. This work was supported by grants from the National Science Foundation and the National Institutes of Health.