The EMBO Journal vol.12 no.1 pp.35-44, 1993

Variation of half-site organization and DNA looping by AraC protein

John H.Carra and Robert F.Schleif

Biology Department, Johns Hopkins University, 34th and Charles Streets, Baltimore, MA 21218, USA

Communicated by H.Bujard

The dimeric AraC protein of *Escherichia coli* binds specifically to DNA sequences upstream of promoters whose transcription is regulated by arabinose. Here we show with affinity measurements, DNase footprinting, dimethyl sulfate premethylation interference and dimethyl sulfate footprinting studies that AraC protein can recognize paired half-sites in direct repeat orientation or inverted repeat orientation. A similar high degree of flexibility was also seen in the ability of the protein in the absence of arabinose to bind tightly and specifically when the separation of its half-sites was increased by 10 or 21 bp. In the presence of arabinose the protein could specifically contact both half-sites of a +10 bp spacing construct but could not contact both in a +21 bp construct. Reduced extensibility of AraC protein in the presence of arabinose provides a simple mechanism for the protein's shift from a non-inducing, DNA looping state to an inducing, non-looping state that contacts two adjacent half-sites at the $araP_{BAD}$ promoter.

Key words: domains/direct repeat/flexible linker/transcription activator

Introduction

In *Escherichia coli* the products of the *araBAD* operon catabolize L-arabinose (Schleif, 1987). Transcriptional regulation of the promoter p_{BAD} is mediated by the AraC protein, a dimer of 65 000 dalton mol. wt (Wilcox and Meuris, 1976; Steffen and Schleif, 1977), which positively activates transcription from p_{BAD} in the presence of the inducer arabinose (Englesberg, 1961; Englesberg *et al.*, 1965, 1969; Greenblatt and Schleif, 1971). In the absence of arabinose, an AraC protein dimer simultaneously binds to the two half-sites, $araI_1$ and $araO_2$, which are 210 bp apart, thereby forming a DNA loop (Figure 1) (Ogden *et al.*, 1980; Hahn and Schleif, 1983; Dunn and Schleif, 1984; Dunn *et al.*, 1984; Martin *et al.*, 1986; Hamilton and Lee, 1988; Lobell and Schleif, 1990). The addition of arabinose causes the loop to open and induction of transcription follows. DNA looping was first demonstrated in the arabinose system (Dunn *et al.*, 1984; Martin *et al.*, 1986). It is now known to be a widespread phenomenon involved in transcriptional regulation in many organisms as well as in DNA recombination and in the control of DNA replication (Schleif, 1987, 1992). DNA looping is likely to be regulated in some systems; one result of the work described here is a mechanism for such regulation.

In the *araBAD* operon, DNA loops can form with many different integral numbers of helical turns of DNA, from one to over 20, between the $araO_2$ and $araI_1$ half-sites (Lee and Schleif, 1989). This suggests that the AraC protein is highly flexible. In addition, flexibility appears necessary to allow a shift in the AraC protein–DNA complexes between the looped and unlooped states (Lobell and Schleif, 1990). In this paper we explore the extent of this flexibility and address its significance in the regulation of DNA looping.

In vitro studies show that the addition of arabinose opens the I_1-O_2 DNA loop (Lobell and Schleif, 1990) and that the AraC protein dimer changes the location of its protein–DNA interactions from the well separated I_1 and O_2 half-sites to the adjacent I_1 and I_2 half-sites. This isomerization occurs without dissociation of the protein from the DNA, and does not require the presence of the I_2 half-site. The AraC protein dimer bound at I_1 and I_2 in the presence of arabinose now activates RNA polymerase to transcribe p_{BAD}. Thus the regulatory switch to an inducing state is triggered by arabinose and is mediated by a change in occupancy between the O_2 and I_2 half-sites. One role of the I_1-O_2 DNA loop in this system is apparently to reduce inappropriate occupancy of I_2 and unnecessary transcription of p_{BAD} in the absence of arabinose. One question addressed here is how arabinose triggers the shift from binding to non-adjacent half-sites, which forms a loop, to the binding of adjacent half-sites.

Although previous stoichiometry and contacting studies have shown that it is a dimer of AraC that contacts the *araI* site, they could not determine definitively the symmetry of the sites recognized by AraC protein (Hendrickson and Schleif, 1985). Within the region of *araI* DNA generally protected from DNase digestion by AraC protein lies a prominent inverted repeat sequence, with each repeat being 13 bp. The pattern of AraC contacts with the DNA phosphate backbone of *araI* also fits the inverted repeat pattern reasonably well (Hendrickson and Schleif, 1985). Alignment, however, of all known AraC protein binding sites suggests that the protein might bind to two direct repeats of 17 bp separated by 4 bp (Lee *et al.*, 1987; Martin, 1987). Missing contact footprinting of two alanine substituted AraC mutants provided strong, but not absolutely convincing evidence that AraC protein recognizes the 17 bp sequence in a direct repeat orientation (Brunelle and Schleif, 1989). Unfortunately, the locations of the missing contacts observed do not totally resolve the symmetry issue. The locations of the missing contacts fit approximately both a 13 bp inverted repeat, and a 17 bp direct repeat (Figure 1).

Results presented in this paper from the study of an artificial binding site with two I_1 half-sites prove that the relevant symmetry of the *araI* site is a direct repeat. In addition, these studies show that a dimer of AraC protein can also simultaneously tightly bind to two I_1 half-sites in inverted repeat orientation. The AraC protein dimer can contact two directly repeated half-sites when their separation is increased by 10 or 21 bp, but when arabinose is added,

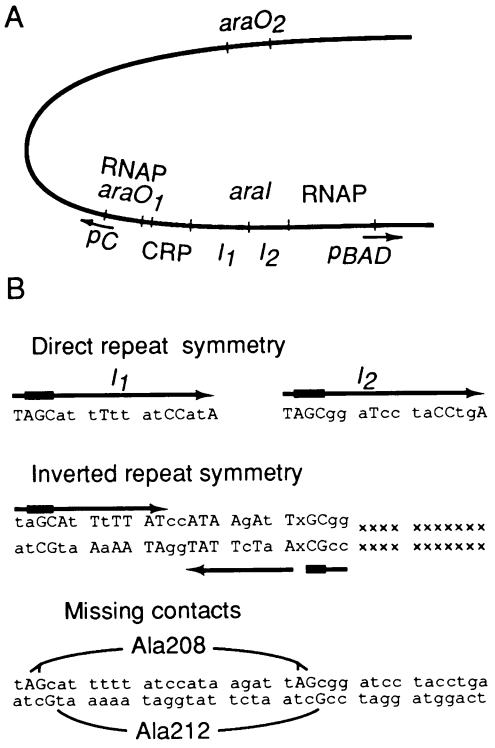

Fig. 1. (**A**) The binding sites and promoters of the regulatory region of the *araBAD* operon. The 210 bp separating the $araO_2$ and *araI* site are shown bent into a looped state. (**B**) The *araI* site. Shown above is a decomposition of the *araI* sequence into direct and inverted repeat symmetry elements. The right half of the inverted symmetry element is interrupted. The thicker portions of the arrows indicate that the nucleotides of the binding site identified by missing contact probing to lie near the ends of the arrows for both symmetry possibilities. Bases which fit the direct or inverted repeat consensus, respectively are upper case. Below are the locations of protein–DNA contacts lost by two alanine-substituted AraC mutants (Brunelle and Schleif, 1989).

simultaneous binding of both protein subunits across 21 additional base pairs is strongly reduced. These results show that arabinose can drive the relocation of AraC protein from looping to nonlooping by restricting the protein to binding to closely spaced half-sites.

Results

Strong, specific binding to a direct repeat binding site

To obtain more information on the relevant symmetry of the *araI* site we synthesized an AraC protein binding site with a direct repeat of the left half-site of *araI* (Figure 1). This is a close fit to the consensus half-sites of all the known AraC binding sites (Martin, 1987). This part of the *araI* site appears to bind particularly tightly because various footprinting and contact probing methods reveal strong protein–DNA contacts in this region (Ogden *et al.*, 1980; Hendrickson and Schleif, 1984; Brunelle and Schleif, 1989). We made the I_1-I_1 site whose sequence is given in Materials and methods and characterized its interactions with AraC protein. In summary, gel retardation assays showed that AraC protein binds the I_1-I_1 site much more tightly than the natural *araI* site, and DNase I footprinting, dimethyl sulfate (DMS) footprinting and pre-methylation interference assays showed that the pattern of AraC -I_1-I_1 contacts fits a direct repeat symmetry.

The equilibrium binding constant and dissociation rate of the AraC-I_1-I_1 complex were measured with the DNA

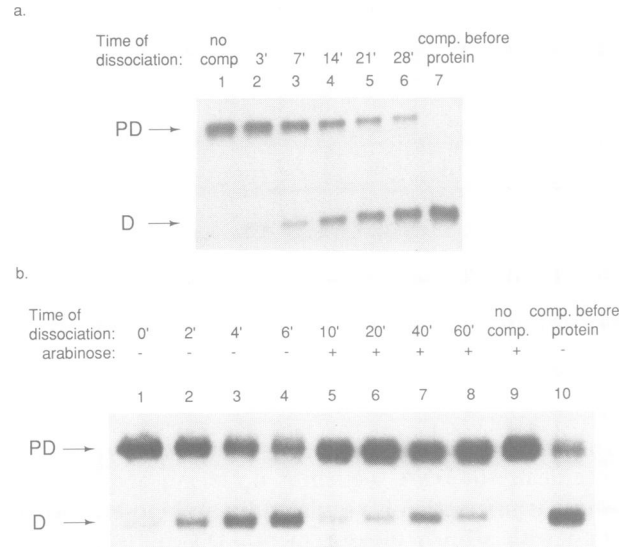

Fig. 2. (**a**) DNA mobility shift measuring the rate of protein dissociation from the I_1-I_1 site. After formation of protein–DNA complexes on the radioactively labeled I_1-I_1 site, an ~100-fold excess of unlabeled I_1-I_1 competitor DNA was added and dissociation was allowed to proceed for varying lengths of time as indicated. Dissociation was stopped by loading all of the reaction on a retardation assay gel. The binding buffer contained 150 mM KCl and no arabinose. Lane 1 had no specific competitor added, and shows that all of the labeled DNA was bound by protein before the addition of competitor. Lane 7 had the I_1-I_1 competitor added before the AraC protein, and shows that the competitor was present in large excess over the labeled DNA, making dissociation from the labeled DNA effectively irreversible. (**b**) DNA mobility shift assay measuring the rate of protein dissociation from the I_1-*inv*-I_1 site. This was performed as in panel a except that the binding buffer contained 50 mM KCl and arabinose as indicated. Lane 9 had no specific competitor added and lane 10 had the I_1-I_1 competitor added before the AraC protein.

Table I. Affinities of binding sites[a]

	$t_{1/2}$ of dissociation		Apparent K_{eq} of binding	
Site	−ara	+ara	−ara	+ara
araI (I_1-I_2) (50 mM KCl)	3 min	100 min	2×10^{11} M^{-1}	8×10^{12} M^{-1}
araI (I_1-I_2) (150 mM KCl)	<20 s	3 min	5×10^{9} M^{-1}	2×10^{11} M^{-1}
I_1-I_1 (50 mM KCl)	>200 min	>200 min		
I_1-I_1 (150 mM KCl)	15 min	>200 min	5×10^{11} M^{-1}	$\geq2\times10^{13}$ M^{-1}
I_1-*inv*-I_1 (50 mM KCl)	10 min	>200 min	6×10^{12} M^{-1}	$\geq3\times10^{13}$ M^{-1}
I_1-5-I_1 (50 mM KCl)	<20 s	10 min	2×10^{10} M^{-1}	8×10^{11} M^{-1}
I_1-10-I_1 (50 mM KCl)	30 min	≥200 min	2×10^{12} M^{-1}	$\geq8\times10^{13}M^{-1}$
I_1-21-I_1 (50 mM KCl)	10 min	30 min	7×10^{11} M^{-1}	2×10^{12} M^{-1}

[a]Possible errors on both rates of dissociation and equilibrium binding constants are likely to be <2-fold.

mobility shift assay. As AraC protein concentration or time of protein–DNA complex dissociation was varied in the assay, a single transition was observed from bound to unbound species with no evidence of stable intermediates (Figure 2a). At 150 mM KCl, dissociation occurs from the I_1-I_1 site with a $t_{1/2}$ of ~15 min in the absence of arabinose.

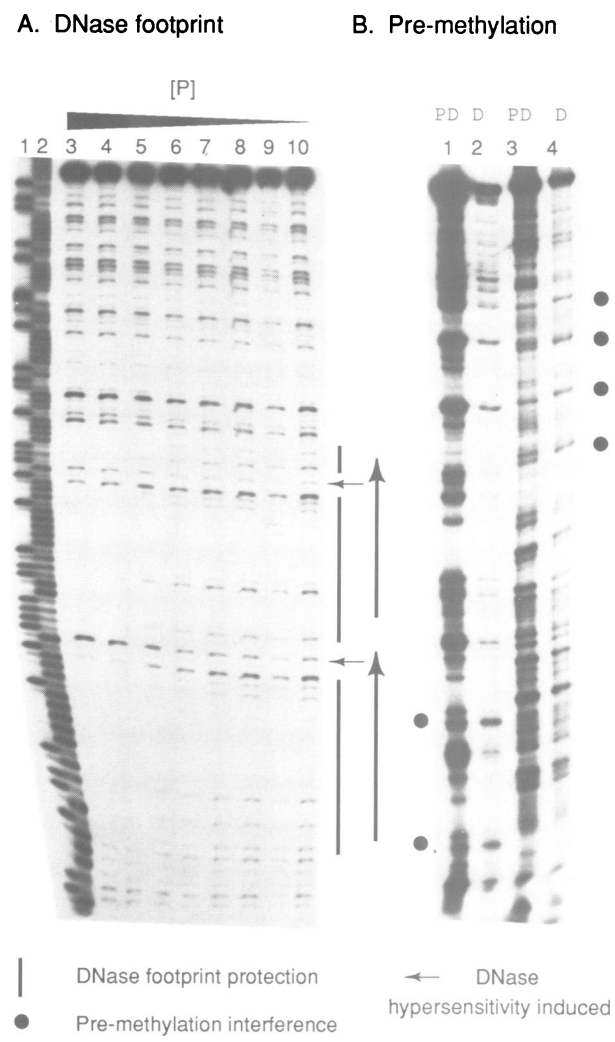

Fig. 3. (**A**) DNase footprint of the *araI*$_1$-I_1 site. Horizontal arrows indicate locations of AraC protein-induced DNase protections. Vertical arrows delimit the two I_1 half-sites. The top strand of the I_1-I_1 site shown in Figure 1 was 5′ end-labeled. Lanes 1 and 2 contain Maxam–Gilbert G+A and C+T sequencing reactions, respectively. Approximate amounts of dimeric AraC in pmol added to the lanes were lane 3, 1.5; lane 4, 0.75; lane 5, 0.37; lane 6, 0.18; lane 7, 0.09; lane 8, 0.05; lane 9, 0.02; and lane 10, 0. (**B**) Pre-methylation interference assay of the I_1-I_1 site. Circles indicate guanine residues where pre-methylation interfered with AraC protein binding. Lanes 1 and 2 show contacts to the top strand of the I_1-I_1 site, and lanes 3 and 4 to the bottom strand.

In the presence of the sugar, dissociation is very slow, with a $t_{½}$ of >200 min. For comparison, at 150 mM KCl AraC dissociates from the natural *araI* site with a $t_{½}$ of a few seconds in the absence of arabinose and 3 min in its presence (Hendrickson and Schleif, 1984). The AraC–I_1-I_1 complex therefore dissociates on the order of 100 times more slowly than the AraC–*araI* complex. Measurement of the equilibrium constant of binding of the I_1-I_1 site yields a similar result of ~100-fold greater stability (Table I). The high affinity of the I_1-I_1 site for AraC protein is consistent with the hypothesis that AraC protein recognizes a 17 bp direct repeat.

DNase footprinting (Figure 3A) was performed to determine the locations and tightness of the protein–DNA interactions in the AraC–I_1-I_1 complexes. At each half-site, AraC protein both protects from DNase digestion and induces internal DNase hypersensitivities. The extent of the protection and the identical locations of the hypersensitivities shows that a subunit of AraC is bound at both half-sites of I_1-I_1. As the protein concentration in these reactions is decreased, footprinting at both half-sites and binding in the DNA mobility shift assay are lost in concert, showing that the complex observed in the DNA mobility shift assay has both half-sites occupied by protein. These results indicate that on the I_1-I_1 site AraC binds either to both half-sites simultaneously or to neither. Binding of protein monomers or of only one subunit of a dimer is not observed. The observation that DNase protection is complete at both half-sites of I_1-I_1 supports the conclusion that both are recognized specifically.

Figure 3B shows the results of a pre-methylation interference assay on the I_1-I_1 site. The results, summarized in Figure 4, show direct repeat symmetry in the patterns of DNase digestion and premethylation interference. The interactions of the protein with the I_1-I_1 site are the same in each 17 bp half-site and identical to those observed on the I_1 sequence of the natural *araI* site (Hendrickson and Schleif, 1985), demonstrating that AraC recognizes a direct repeat as opposed to the previously hypothesized inverted repeat (Hendrickson and Schleif, 1985).

AraC does not oligomerize indefinitely on DNA

AraC is predominantly a dimer in solution (Wilcox and Meuris, 1976) and also binds to *araI* as a dimer (Hendrickson and Schleif, 1985). Nonetheless, the conclusion that AraC protein recognizes a direct repeat sequence raises the possibility that AraC protein could polymerize indefinitely on DNA containing multiple direct repeats of its binding half-sites. This possibility was tested by synthesizing a binding site consisting of four I_1 half-sites tandemly arranged head-to-tail and spaced to lie on the same face of the DNA helix. Figure 5 shows a DNA mobility shift assay of AraC protein–DNA complexes formed on the I_1 site. Two different protein–DNA complexes are observed. The more slowly migrating complex is presumably two dimeric units of AraC protein per molecule of DNA, denoted by P_2D. This complex predominates at higher protein concentrations. Kinetic experiments (data not shown) show that this complex dissociates through the more rapidly migrating complex, PD, as an intermediate. Both bound species dissociate at a rate similar to that seen from the I_1-I_1 site complex. This indicates that in each case, a protein dimer dissociates.

DNase footprinting and pre-methylation interference experiments confirm that the P_2D complex has all four half-sites of the $4I_1$ site completely occupied, while the PD complex consists of a mixture of isomers each having two adjacent half-sites occupied. In both the presence and absence of arabinose, AraC protein binds to DNA in units of dimers. Evidence for trimers was not seen. Additionally, no cooperativity between dimers was observed in the equilibrium binding or dissociation rate. Similar studies of a $3I_1$ site also reveal AraC protein binding only in units of dimers. These data also show no evidence for interaction between a dimer and monomer or between two dimers.

AraC can also bind specifically to an inverted repeat

The finding that AraC protein recognizes the *araI* site as a direct repeat raises the question of how the symmetry of the DNA half-sites relates to that of the protein. Most prokaryotic DNA binding proteins of known structure (Aggarwall and Harrison, 1990) oligomerize via isologous,

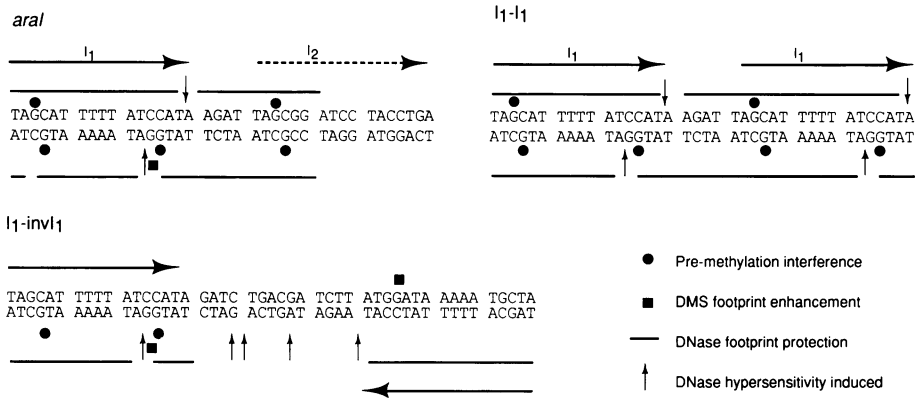

Fig. 4. Protein–DNA contacts to the *araI*, I_1-I_1 and I_1-*inv*-I_1 sites. Large horizontal arrows show the locations of half-sites. Small, vertical arrows indicate locations of the AraC protein-induced DNase hypersensitive sites. Horizontal bars show areas of AraC protein-induced DNase protections. Circles indicate guanine residues where pre-methylation interfered with AraC protein binding. Squares indicate guanines which are hypermethylated by DMS in the presence of AraC protein.

head-to-head interactions and bind to inverted repeat half-sites. These facts, plus the precedent of separable domains in other regulatory proteins (Pabo *et al.*, 1979; Smith and Johnson, 1992; Weiss *et al.*, 1983) raise the possibility that a dimerization domain and DNA binding domain of AraC are connected by a particularly flexible linker. Some indication already exists that AraC might possess functionally separable domains. Protease treatment of the protein leaves a resistant N-terminal two-thirds of the sequence. This remains dimeric, but shows no significant DNA binding ability (R.F.Schleif, unpublished). Also, *in vivo* expression of the C-terminal 40% of the protein may be capable of weakly activating transcription of the $arap_{BAD}$ promoter (Menon and Lee, 1990). We might, therefore, expect flexibly linked DNA binding domains of the AraC protein to be capable of lying on the DNA in either head-to-tail or head-to-head orientation while binding to direct or inverted repeats, respectively.

The hypothesis that the AraC dimer can specifically contact two half-sites in inverted orientation was tested directly. Thirteen base pairs of spacer were inserted between inverted half-sites so that they remained on the same face of the DNA. The resulting 47 bp inverted repeat sequence was placed in the same sequence context, flanked by the same 44 and 52 bp as were used in all the studies reported here. The $t_{½}$ of AraC dissociation from I_1-*inv*-I_1 is 10 min in buffer containing 50 mM KCl and >200 min when arabinose is added to the buffer (Table I, Figure 2b). Dissociation from I_1-*inv*-I_1 is thus faster than from I_1-I_1, (Table I) but slower than from I_1-I_2. The reduction in electrophoretic mobility caused by the binding of AraC protein is approximately the same as with I_1-I_2, indicating that it is dimer of AraC protein which binds to I_1-*inv*-I_1, since a dimer is known to bind to I_1-I_2 (Hendrickson and Schleif, 1985).

The DNase footprint of the I_1-*inv*-I_1 site, Figure 6A, shows that in both the presence and absence of arabinose, AraC protein occupies both half-sites in concert. The location of induced DNase hypersensitive bonds within the two half-sites parallels the half-site reorientation. The DNA in the gap between the two half-sites shows prominent induced DNase hypersensitivities, as well as protections. These patterns are consistent with changes in susceptibility to DNase cleavage resulting from alterations in the widths of

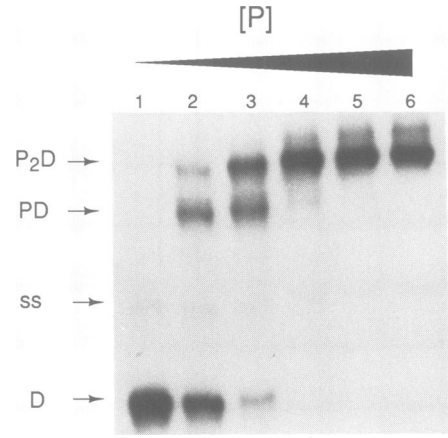

Fig. 5. Gel retardation assay on the $4I_1$ site. AraC protein was titrated against labeled $4I_1$ site. SS indicates single-stranded DNA contaminating the DNA preparation. Approximate amounts of dimeric AraC protein in pmoles added to the reactions were lane 1, 0; lane 2, 0.004; lane 3, 0.008; lane 4, 0.015; lane 5, 0.030; and lane 6, 0.060.

the minor grooves generated by bending between the two half-sites (Drew and Travers, 1984). The location of induced DNase hypersensitivities on the I_1-*inv*-I_1 complex is different by a few base pairs in the presence and absence of arabinose (Figure 6A). This change suggests that the arabinose-bound and unbound complexes have slightly different geometries due to a conformational change in the AraC protein. Figure 6B and C show the results of application of pre-methylation interference techniques and DMS footprinting to the I_1-*inv*-I_1 site. Although the amounts of radioactivity loaded in the lanes of part C were not equal, the essential point is the enhanced band marked with the square. The results of these two methods and of the DNase footprinting (Figure 4) are consistent with AraC protein specifically recognizing both I_1 half-sites in the inverted orientation.

AraC cannot reach halfway around the DNA helix

The existence of a flexible connection between the dimerization and DNA binding domains suggests that AraC protein may not be limited to binding half-sites on the same face of the DNA. If the linker is very long and flexible, the protein may be able to reach part of the way around the

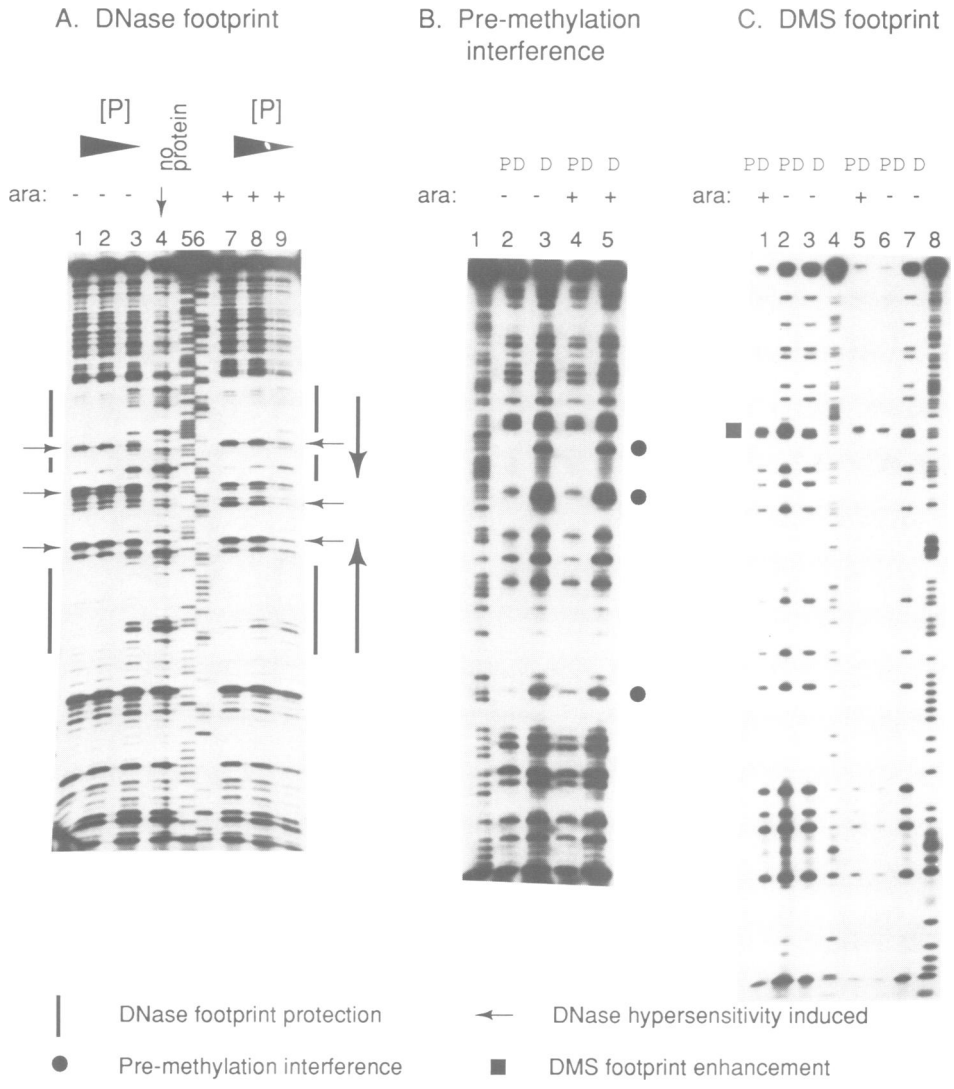

Fig. 6. (**A**) DNase footprint of the I_1-*inv*-I_1 site. Horizontal arrows indicate locations of AraC protein-induced DNase hypersensitive sites. Vertical bars show areas of AraC protein-induced DNase protection. Vertical arrows delimit the two I_1 half-sites. The bottom strand of the I_1-*inv*-I_1 site shown in Figure 4 was 5′ end-labeled. Lanes 5 and 6 contain Maxam−Gilbert G+A and C+T sequencing reactions, respectively. Approximate amounts of dimeric AraC protein in pmol added were lane1, 3; lane 2, 1; lane 3, 0.3; lane 4, 0; lane 7, 3; lane 8, 1; and lane 9, 0.3. (**B**) Pre-methylation interference assay on the I_1-*inv*-I_1 site. Circles indicate guanine residues where pre-methylation interfered with AraC protein binding. The bottom strand of the I_1-*inv*-I_1 site was 5′ end-labeled. Lane 1 shows a Maxam−Gilbert G+A sequencing reaction. (**C**) DMS footprinting assay on the I_1-*inv*-I_1 site. Squares show guanines which are hypermethylated by DMS in the presence of AraC protein. Contacts to the bottom strand of I_1-*inv*-I_1 site are shown. Lanes 3 and 7 show DMS-reacted DNA without selection for AraC protein binding. Lanes 4 and 8 show a Maxam−Gilbert G+A sequencing reaction. Lanes 1−4 are bottom strand, 5−8, top strand.

circumference of the DNA, allowing a dimer of the protein to contact simultaneously two half-sites that are out of helical phase. This possibility was explored with a variant of the I_1-I_1 site, I_1-5-I_1, which contains an additional 5 bp of DNA inserted between the half-sites. Using the DNA mobility shift assay, premodification of the DNA with hydroxyl radical treatment, and DNase footprinting, we found that AraC protein binds to the half-sites of I_1-5-I_1 independently in both the presence and absence of arabinose (data not shown). In other words, one protein dimer could not simultaneously bind both half-sites.

Using the I_1-5-I_1 site, two different weak protein−DNA complexes were observed with the gel retardation assay. One apparently contains a single protein dimer per DNA molecule, PD, and the other contains two, P_2D. As the concentration of AraC was varied, the ratios of the D, PD and P_2D species varied as expected for independent binding of protein dimers. Furthermore, the P_2D complex dissociated through the PD complex as an intermediate. Both the PD and P_2D complexes had higher dissociation rates than the *AraC*−I_1-I_1 complex. The result is also consistent with the protein dimer specifically contacting one I_1 half-site and non-specifically contacting other DNA. In conclusion, the AraC protein dimer cannot simultaneously contact two half-sites on opposites faces of the DNA.

Arabinose decreases the effective arm length of AraC protein

Arabinose destabilizes AraC protein binding to the I_1-O_2 looped complex, but stabilizes binding to the I_1-I_2 or I_1-I_1 sites. If we hypothesize that it is the separation between the half-sites that determines this response, how far must the half-sites be separated in order for arabinose to destabilize the interaction? To answer this question, a site with the

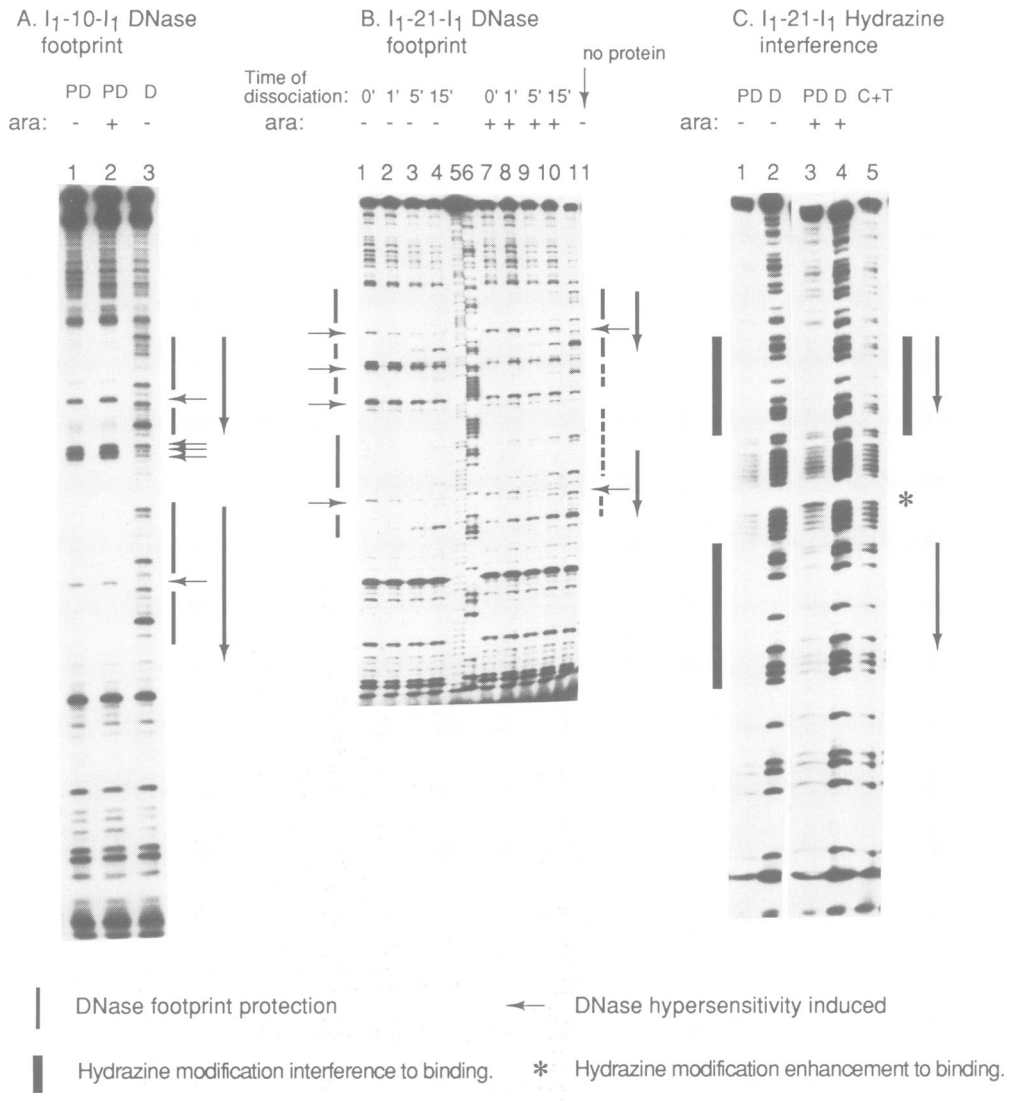

Fig. 7. (**A**) DNase footprinting of I_1-*10*-I_1 site. Horizontal arrows indicate locations of AraC protein-induced DNase hypersensitivities. Vertical bars show areas of AraC protein-induced DNase protections. Vertical arrows delimit the two I_1 half-sites. The bottom strand of the I_1-*10*-I_1 site 1 was 5′ end-labeled. (**B**) DNase footprint of the I_1-*21*-I_1 site. Horizontal arrows indicate locations of AraC protein-induced DNase hypersensitive cleavages. Vertical bars show areas of AraC protein-induced DNase protection. Dashed vertical bars show areas of partial DNase protection. Vertical arrows delimit the two I_1 half-sites. The bottom strand of the I_1-*21*-I_1 shown in Figure 8 was 5′ end-labeled. Lanes 5 and 6 contain Maxam–Gilbert G+A and C+T sequencing reactions, respectively. Protein–DNA complexes were formed and then allowed to dissociate in the presence of an excess of unlabeled I_1-I_1 competitor over time. One-twentieth of the reaction was used on a DNA mobility shift assay to determine how much of the DNA was bound by the protein. Almost all of the labeled DNA in lanes 2 and 8 was present in the PD complex. (**C**) Hydrazine modification interference assay on the I_1-*21*-I_1 site. Vertical bars show areas where pre-modification of the DNA with hydrazine interfered with AraC protein binding. An asterisk indicates a position where hydrazine modification enhanced AraC binding.

normal four base separation, I_1-I_1, as well as sites with an additional 10, 21 32, and 43 bp separating the half-sites, I_1-*10*-I_1, I_1-*21*-I_1, I_1-*32*-I_1 and I_1-*43*-I_1, were tested. We attempted to facilitate looping between the half-sites by inserting tracts of bent DNA in the spacer region between them. The naturally bent sequence A_5 (Wu and Crothers, 1984), when placed between half-sites may be expected to facilitate the formation of a looped geometry in the DNA. In each site one A_5 sequence was inserted per turn of DNA in the spacer, positioned so as to promote a bend in the DNA in the direction of looping. The I_1-*inv*-I_1 site discussed above lacks the A_5 sequence, yet still forms a sort of loop, indicating that the bent DNA may be beneficial to looping on linear DNA but is not essential.

DNA mobility shift assays, DNase footprinting and hydrazine pre-modification interference experiments done in the presence and absence of arabinose show that the dimeric AraC protein simultaneously contacts both half-sites on the I_1-I_1 and I_1-*10*-I_1 sites in the presence and absence of arabinose, but can contact both half-sites of I_1-*21*-I_1 only in the absence of arabinose. Table I shows the apparent equilibrium constants as measured by competition assay, and the half-times of dissociation of AraC protein from the natural I_1-I_2 site, the I_1-I_1 site, and the I_1-*10*-I_1 and I_1-*21*-I_1 sites. The relative equilibrium constants of AraC protein binding to these sites parallel their dissociation rates under all conditions assayed. It was necessary to use 150 mM KCl when measuring the I_1-I_1 site rather than 50 mM KCl, which was used for the other sites, because the AraC–I_1-I_1 complex does not dissociate measurably at the lower salt concentration. The I_1-I_1, I_1-*10*-I_1 and I_1-*21*-I_1 sites all bind AraC protein more tightly than the natural *araI* site, I_1-I_2,

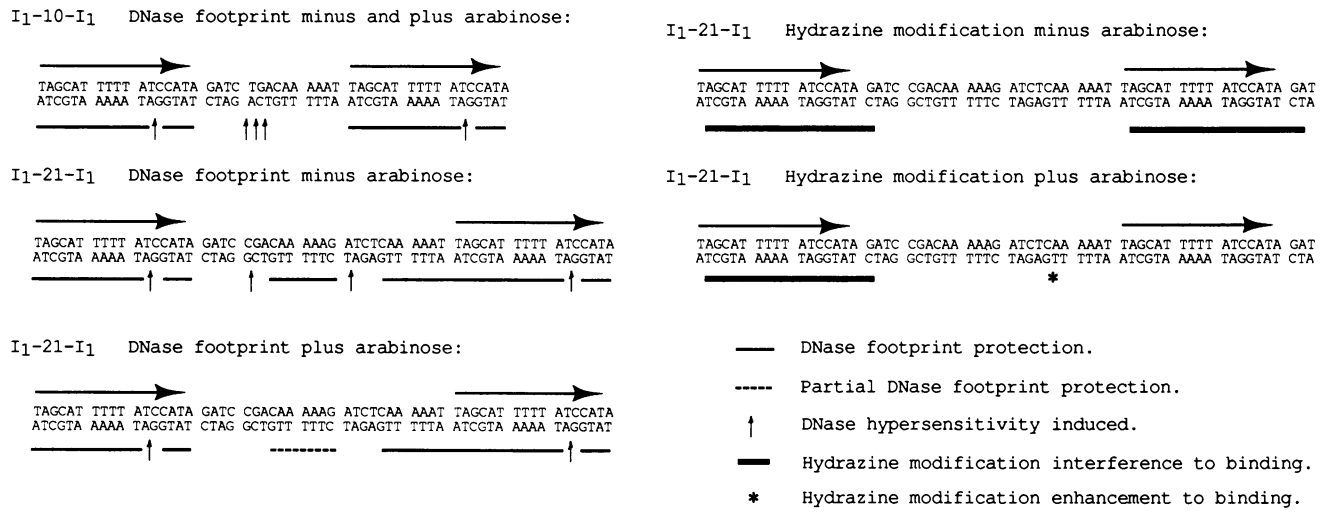

Fig. 8. Results from DNase footprinting and hydrazine modification of T residues in the I_1-*10*-I_1 and I_1-*21*-I_1 sites. Large horizontal arrows show the locations of half-sites. Small vertical arrows indicate locations of AraC protein-induced DNase hypersensitive sites. Horizontal bars show areas of AraC protein-induced DNase protection. Dashed horizontal bars show areas of partial DNase protection. Results for the bottom strand only of each site are shown. The horizontal bars show areas of hydrazine modification interference. Thick vertical bars indicate positions where modification of pyramidines with hydrazine interfered with AraC protein binding. A star indicates a position where hydrazine modification enhanced AraC protein binding. Lane 5 shows a hydrazine-reacted DNA without selection for AraC protein binding. The bottom strand of the I_1-*21*-I_1 site shown in this figure was 5′ end-labeled.

in the absence of arabinose. Arabinose strengthens binding to the I_1-I_2, I_1-I_1 and I_1-*10*-I_1 sites by ≥40-fold. Binding to the I_1-*21*-I_1 site, however, is stabilized by arabinose only 3-fold, and the complex now dissociates more quickly than the AraC−I_1-I_2 complex.

DNase footprinting, (Figure 3, summarized in Figure 4) shows that AraC protein completely protects both half-sites of the I_1-I_1 site from DNase digestion except for two internal points of hypersensitivity. These results are obtained in both the presence and absence of arabinose. The footprint of the I_1-*10*-I_1 site also indicates complete binding to both half-sites with and without the sugar (Figures 7 and 8). The pattern of DNase digestion of the complex with the I_1-*10*-I_1 site shows positions of both enhanced and reduced digestion rate between half-sites. DNase sensitivity patterns of this sort indicate the presence of looping (Hochschild and Ptashne, 1986). Although the two binding half-sites in these experiments and others reported here are identical, the footprinting experiments display minor differences in the two halves. These differences probably result from the very fact that in the case of directly repeated half-sites, the two DNA-contacting domains cannot be identical as they can be in the case of inverted repeats. Both the DNA and protein possess an inherent asymmetry, and this is probably reflected in the minor footprinting differences.

The I_1-*21*-I_1 site yields a DNase footprint similar to that of the I_1-*10*-I_1 site in the absence of arabinose. It shows complete occupancy of both half-sites and induced hypersensitivities in the spacer region. In the presence of arabinose, however, the I_1-*21*-I_1 site no longer shows complete occupancy of both half-sites, or the hypersensitivities in the spacer. Simultaneous application of the band-shift assay and DNase footprinting techniques to the reaction mixtures of Figure 7B revealed that while all of the labeled DNA in lanes 2 and 8 was present in protein−DNA complexes, DNase protection at the right half-site (Figure 8) was incomplete in the presence of arabinose. The complex on the I_1-*21*-I_1 site therefore has both half-sites occupied by AraC in the absence of arabinose, but not in its presence. In other words an arabinose-bound AraC protein dimer contacts only one of the two I_1-*21*-I_1 half-sites at a time.

A change in DNase footprint pattern upon the addition of arabinose similar to that of the I_1-*21*-I_1 site was also observed with the more widely spaced I_1-*32*-I_1 and I_1-*43*-I_1 sites. These two sites form weak loops with half times of dissociation of <5 min at 50 mM KCl. The stability of the complexes formed in the absence of arabinose decreases progressively with the increasing length of the spacer DNA. These results show that simultaneous binding by AraC protein to two well separated sites is difficult on linear DNA. On supercoiled DNA such binding is much easier and readily occurs (Lee and Schleif, 1989; Lobell and Schleif, 1990).

Hydrazine modification interference assays were done to confirm the identification of the complexes formed by the I_1-*21*-I_1 site. Labeled DNA of the binding site was partially reacted with hydrazine and used in a gel retardation experiment analogous to pre-methylation interference experiments. Those molecules with disruptions in bases that are necessary for AraC binding will be bound less tightly by the protein and will be found preferentially in the free DNA species on the gel. This DNA was extracted, cleaved with piperidine and run on a DNA sequencing gel. Any position where disruption interferes with recognition of the DNA by AraC will be enhanced in intensity relative to non-contacted positions. The converse is found for DNA extracted from the AraC−DNA complex. Figures 7C and 8 show that, and quantification of the relevant bands with a Phosphoimager confirms that, in the absence of arabinose, hydrazine modification at either half-site of the I_1-*21*-I_1 site interferes with AraC protein binding. When arabinose is present, however, interferences are predominantly observed only at one half-site. In addition an enhancement to binding is now observed at one point in the spacer region. We interpret these results as meaning that arabinose has caused AraC protein to shift from contacting both I_1-*21*-I_1 half-sites simultaneously to contacting one half-site and the adjacent spacer region.

The results of the hydrazine interference technique and

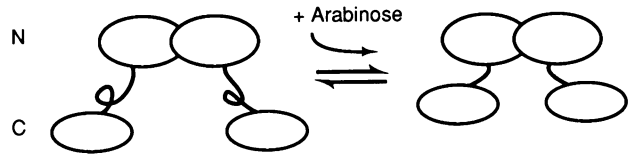

Fig. 9. Hypothetical conformational change in AraC protein induced by arabinose addition.

pre-methylation interference technique (data not shown) applied to the I_1-21-I_1 site are consistent with the hypothesis that in the presence of arabinose interactions of AraC protein with two half-sites separated by ≥21 bp are not possible. Instead in the presence of arabinose AraC protein prefers to bind one I_1 half-site and the best sequence nearby, which in this case is within the spacer region. The enhancement to binding observed at one position in the spacer probably results from the disruption by hydrazine of a base whose presence impedes binding to the spacer DNA.

The 3-fold stabilization by arabinose of binding to the I_1-21-I_1 site can be understood as a combination of the stabilization by the sugar of AraC protein's proximal interactions with one I_1 half-site and the spacer region, and destabilization of distal interactions across two I_1 half-sites.

Discussion

We have shown here that the dimeric AraC protein possesses a surprising degree of flexibility in that it can specifically contact its half-sites when they are in direct repeat orientation or inverted repeat orientation. Further, it can specifically contact these half-sites when their separation is increased by 10 or 21 bp. Finally, the presence of arabinose decreases the apparent extensibility of AraC protein, preventing simultaneous binding to half-sites separated by ≥21 bp. A simple hypothesis consistent with these observations is that AraC is composed of a dimerization domain loosely connected to a DNA binding domain, and that the connection between the two domains permits both extension and rotation. In the presence of arabinose the connection between the two domains is tightened.

The effect of arabinose upon AraC protein's interactions with the sites having larger spacings mimics the loop opening observed *in vivo* with the natural O_2 and I_1 half-sites. That is, the presence of arabinose weakens binding to widely separated half-sites and strengthens binding to adjacent sites. It follows that the differential effect of arabinose which is observed on AraC protein's binding to the alternative half-sites O_2 and I_2 is determined by the half-sites' spatial location, not their specific sequences. That is, looping of AraC between $araI_1$ and $araO_2$ is dramatically reduced by the addition of arabinose, independent of the presence of $araI_2$ (Lobell and Schleif, 1990). A mechanism for the regulatory switch of the *araBAD* operon from a non-inducing to an inducing state is a switch from binding to the two best available half-sites within several hundred base pairs in the absence of arabinose to binding to two adjacent half-sites in the presence of arabinose.

A hypothetical structure for the AraC protein dimer that helps explain the above results is shown in Figure 9. The AraC protein dimer is shown with each subunit having an N-terminal dimerization domain and a C-terminal DNA binding domain. A flexible stretch of amino acids links the two. This linker must be sufficiently extensible to allow the DNA binding domains to contact half-sites that are separated by integral numbers of turns of DNA. The linker must also be sufficiently free to rotate that the DNA binding domains of the protein can bind without very large energy costs to half-sites in direct repeat or inverted repeat orientation. The linker is not long enough, however, to reach halfway around the double helix.

Arabinose shifts AraC protein to a preference for closely separated half-sites. This occurs *in vivo* when the protein shifts from its non-inducing state of looping between $araI_1$ and $araO_2$ to its inducing state of binding $araI_1$ and $araI_2$. Taking into consideration all the experimental results, we hypothesize that arabinose tightens the connection between the dimerization and DNA binding domains. The arabinose-bound structure of the protein must still allow some flexibility, including rotational freedom, to accommodate our findings that arabinose stabilizes binding to half-sites separated by zero or one turn of DNA, and in direct or inverted repeat arrangement. The linker of the arabinose-bound protein is too short, however, to permit the protein to reach across two half-sites in a distal organization, such as with the I_1-21-I_1 site or the natural I_1 and O_2 half-sites. The two different protein conformations of Figure 9 can be considered to be the repressing and inducing conformations of AraC protein postulated long ago (Englesberg *et al.*, 1969).

The above-mentioned hypothesis for the arabinose induced conformational change can explain both why the sugar destabilizes distally bound species, loops, as well as why it stabilizes proximally bound species like the natural AraC−I_1-I_2 complex or the AraC−I_1-I_1 complex. A protein dimer with subunit domains connected more closely in a conformation well positioned for binding adjacent half-sites will bind more tightly to adjacent half-sites than a protein in which these domains are more freely connected.

Why then *in vivo* in the absence of arabinose does AraC protein prefer to loop between $araI_1$ and $araO_2$ instead of binding $araI_1$ and $araI_2$? A simple hypothesis is that in this case the two DNA binding domains of AraC choose to bind to the best half-sites available. Although the affinity for the individual half-sites has not been determined, we hypothesize that $araO_2$ binds AraC more tightly than $araI_2$. On supercoiled DNA *in vitro* the looped complex between I_1 and O_2 in the absence of arabinose is very stable, dissociating at a rate comparable to that of a complex with the I_1-I_1 site on linear DNA (Lobell and Schleif, 1990).

Other transcription factors both of prokaryotes and eukaryotes are likely to share certain characteristics of this model for AraC protein structure. AraC protein is a member of a family of more than 20 regulatory proteins possessing significant sequence similarity. The AraC binding sites that are well characterized all possess directly repeated half-sites (Schleif, 1987). Curiously, the one known binding site for RhaR, one of the closest relatives to AraC, is an inverted repeat similar in arrangement to the I_1-*inv*-I_1 site (Tobin and Schleif, 1990). The symmetry of the binding sites and the rotational freedom of the DNA binding domains of the other members of the AraC family remains to be demonstrated.

In yeast, the Matα2 protein has been shown to bind to pairs of half-sites separated by various spacings and in direct or inverted repeat orientations. Consequently, it too has been suggested to contain a flexible linker between dimerization and DNA binding domains (Smith and Johnson, 1992). In conjunction with the MCM1 protein, Matα2 protein can

form a larger cooperative structure containing both proteins, but only when its half-sites are in their natural orientation and 2.5 turns apart. MCM1 protein apparently restricts the DNA binding site specifically of Matα2. This restriction can be considered to be analogous to the effects of arabinose on AraC protein's DNA binding function.

A similar structural model of two loosely connected domains has also been proposed for the GABP transcription factor on the basis of its recognition of directly repeated half-sites and the prediction that it dimerizes by face-to-face isologous interactions (Thompson *et al.*, 1991). Lambda *cI* repressor also has two domains joined by a flexible linker, although the degree of flexibility conferred by this amino acid sequence may not be as great as that proposed for AraC, Matα2 and GABP (Pabo *et al.*, 1979, Weiss *et al.*, 1983; Jordan and Pabo, 1988). Cooperative binding of a lambda *cI* repressor dimer to variably spaced or orientated half-sites has not been reported. Lack of tolerance to variation of spacing and orientation may reflect a lack of great flexibility between domains of this protein or may be a consequence of a need for subunit−subunit contacts between adjacent DNA binding domains (Jordan and Pabo, 1988). Precedent for flexible linkers in proteins can also be found in myosin (Harvey and Cheung, 1982; Curry and Krause, 1991), immunoglobulins (Sosnick *et al.*, 1992) and the ribosomal protein L7/L12 of *Escherichia coli*. L7/L12 is a dimeric protein and contains a highly mobile 13 amino acid hinge region between two structured domains (Bushuev *et al.*, 1989).

The structure of the DNA binding domain of GAL4 (Marmorstein *et al.*, 1992) reveals that this protein contains a 9 amino acid linker between its DNA recognition and dimerization domains. The related protein PPR1 recognizes the same sequence elements as GAL4, but at a spacing of 6 bp between them rather than 11. The DNA binding characteristics of chimeric PPR1/GAL4 proteins (Corton and Johnson, 1989) indicate that it is the linker region of these proteins which determines specificity for a particular spacing.

From the studies reported here as well as the above-mentioned work, it seems reasonable to state that the orientation and spacing of sequence elements recognized by site-specific regulatory proteins can be as important as the internal sequence of those elements. The ability of a DNA binding transcription factor to bind specifically to sequence elements in particular arrangements is a crucial characteristic of that protein's function. Modulation of this specificity provides another level by which the activity of DNA binding proteins may be regulated.

Materials and methods

Plasmids and constructs

DNA oligonucleotides containing various binding sites for AraC protein were synthesized on an Applied Biosystems 381A DNA Synthesizer, purified by FPLC over a Pharmacia Mono-Q anion exchange column (Cubellis *et al.*, 1985) and cloned between the *Nhe*I and *Bam*HI restriction sites of the plasmid pES51 (Huo *et al.*, 1988), thereby placing the various sites in the former location of *araI* in the $arap_{BAD}$ regulatory region. Ligation, transformation, and sequencing of recombinant clones was accomplished by standard DNA manipulation techniques (Schleif and Wensink, 1981; Maniatis *et al.*, 1982). Oligonucleotide sequences were as follows: *araI*, TAGCATTTTTATCCATAagatTAGCGGATCCTACCTGA; I_1-I_1, TAGCATTTTTATCCATAagatTAGCATTTTTATCCATA; $4I_1$, TAGCATTTTTATCCATAagatTAGCATTTTTATCCATAagatTAGCATTTTTATCCATAagatTAGCATTTTTATCCATA; I_1-*inv*-I_1, TAGCATTTTTATCCATAgatctgacgatctTATGGATAAAAATGCTAATCGTAAAAATAGGTATctagactgctagaATACCTATTTTTACGAT; I_1-5-I_1, TAGCATTTTTATCCATAagatagatcTAGCATTTTTATCCATA; I_1-10-I_1, TAGCATTTTATCCATAgatctgacaaaaatTAGCATTTTTATCCATA; and I_1-21-I_1, TAGCATTTTTATCCATAgatccgacaaaaagatctcaaaaatTAGCATTTTTATCCATA.

DNA mobility shift assays

Generation of labeled DNA. DNA molecules for use in all the different types of AraC protein binding assays performed in this study were created *in vitro* using the polymerase chain reaction (PCR). Alkaline lysis miniprep DNA (Maniatis *et al.*, 1982) of pES51 derivatives containing variant binding sites was used as the template for replication. The sequences of the primers used and their locations of hybridization within the *ara* regulatory region of pES51 are shown below. Primer #1 was used in conjunction with primer #2 to produce DNA molecules with the binding site (the I_1-I_1 site is shown) approximately equidistant from the ends of the molecule. In each PCR one primer was ^{32}P-5′-end-labeled and one was unlabeled. Each PCR contained 1−10 ng plasmid template, 100 ng labeled primer, 150 ng unlabeled primer, 7.5 U AmpliTaq DNA polymerase (Perkin−Elmer), 50 mM KCl, 20 mM Tris−HCl pH 8.3, 1.5 mM $MgCl_2$, 0.01% gelatin (Sigma) and 0.2 mM of each of the four deoxynucleotide triphosphates A, C, G and T (Pharmacia Ultrapure), in a reaction volume of 0.1 ml, overlayed with one drop of mineral oil (Perkin−Elmer). Amplification was performed for 30 cycles with a temperature profile of 1 min at 94°C, 2 min at 55°C and 2 min at 72°C. After the last cycle there was an extra 10 min of extension at 72°C. Oil was removed from the finished reaction by extraction with chloroform. The DNA was purified by two sequential runs through 1 ml G-25 Sephadex spin columns (Maniatis *et al.*, 1982) equilibrated in 10 mM Tris−HCl pH 8.0, 1 mM EDTA, 50 mM KCl. As a control the dissociation rates of AraC protein from the natural *araI* site made by restriction digestion of plasmid DNA and *araI* site made by PCR were measured and found to be identical.

The sequences of the primers were as follows: primer #1, 5′-aagtccacatcgattatttg-3′; primer #2, 5′-acgggtatggagaaacagta-3′; *araI* plus flanking DNA, aagtccacatcgattatttcacggcgtcacactttgctagcccaTAGCATTTTTATCCATAagatTAGCATTTTTATCCATagatcctacctgacgctttttatcgcaactctctctgtttctccatacccgt

Assays. DNA mobility shift assays of AraC protein binding to DNA were performed using purified AraC protein free of arabinose essentially as described by Hendrickson and Schleif (1984). The KCl, labeled DNA and AraC protein concentrations as well as times of incubation were varied to suit the design of each experiment. Binding reactions were incubated at 37°C and all mobility shift runs were at 8°C and 6 V/cm. All reactions contained 0.5 μg of sheared calf thymus DNA per 20 μl volume as a competitor of non-specific binding. The buffer, labeled DNA and non-specific competitor were mixed together before protein addition. Arabinose was added to 66 mM when used. In no case was arabinose added to the gel itself.

Dissociation rate measurements were made as described by Hendrickson and Schleif (1984) except that the competitor of specific binding was unlabeled, PCR-generated I_1-I_1 site in an ~100-fold molar excess over the labeled site.

Equilibrium binding constants were measured by competing increasing amounts of an unlabeled binding site against a constant amount of a labeled site in the presence of barely enough AraC protein to bind all of the labeled DNA. The concentration of the protein was at least 100-fold above the sites' K_D of binding. The ratio of the amount of the two types of DNA molecules at the point of 50% binding of the labeled site was taken as the ratio of their equilibrium constants. The equilibrium constant of binding of the natural *araI* site as measured by Hendrickson and Schleif (1984) was then used as a reference to obtain approximate K_{eq} values. Sites close to each other in their affinities were measured in pairs. Reactions were allowed to equilibrate for at least 10 times the $t_{1/2}$ of dissociation of the more slowly dissociating site except in the case of the measurement of the I_1-I_1 and I_1-10-I_1 sites in the presence of arabinose, where the exceedingly slow dissociation of the complexes made this impossible.

DNase footprinting and other protein−DNA contact techniques

DNase footprinting was done as described by Galas and Schmitz, (1978) and Ogden *et al.* (1980) in 150 μl reaction volumes with the two exceptions that PCR-generated DNA was used and that immediately prior to addition of DNase I, 5 μl of the reaction was removed and applied to a DNA mobility shift assay gel. DMS footprinting and pre-methylation interference assays were done essentially as described by Hendrickson and Schleif (1985).

Pre-modification of labeled binding sites at pyrimidines with hydrazine was done essentially as described by Brunelle and Schleif (1989), with the exception that PCR-generated DNA was used. Binding reactions were

performed with varying amounts of AraC protein in 20 μl volumes in a standard binding buffer (Hendrickson and Schleif, 1984) containing 150 mM KCl and 0.5 μg of calf thymus DNA. Reactions were allowed to equilibrate at 37°C for 70 min before loading on to the gel. DNA was isolated from the gel by electrodialysis (Maniatis *et al.*, 1982), cleaved with piperidine (Hendrickson and Schleif, 1985) and run out on a sequencing-type denaturating polyacrylamide gel.

Acknowledgements

We thank Susan Egan for comments on the manuscript, Thadd Reeder, Silvia Bustos, Becky Eustance, Tom Colonna, Casonya Johnson and Sudha Iyer for discussions and Jeff Withey for technical assistance. This work was supported by NIH grant GM18277.

References

Aggarwall,A.K. and Harrison,S.C. (1990) *Annu. Rev. Biochem.*, **59**, 933–969.
Brunelle,A. and Schleif,R. (1989) *J. Mol. Biol.*, **209**, 607–622.
Bushuev,V., Gudkov,A., Liljas,A. and Sepetov,N. (1989) *J. Biol. Chem.*, **264**, 4498–4505.
Corton,J. and Johnston,S. (1989) *Nature*, **340**, 724–727.
Cubellis,M.V., Marino,G., Mayol,L., Piccialli,G. and Sannia,G. (1985) *J. Chromatogr.*, **329**, 406–414.
Curry,J. and Krause,S. (1991) *Biopolymers*, **31**, 1677–1687.
Drew,H.R. and Travers,A.A. (1984) *Cell*, **37**, 491–502.
Dunn,T. and Schleif,R. (1984) *J. Mol. Biol.*, **180**, 201–204.
Dunn,T., Hahn,S., Ogden,S. and Schleif,R. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5017–5020.
Englesberg,E. (1961) *J. Bacteriol.*, **81**, 996–1006.
Englesberg,E., Irr,J. and Lee,N. (1965) *J. Bacteriol.*, **90**, 946–957.
Englesberg,E., Squires,C. and Meronk,F. (1969) *Proc. Natl. Acad. Sci. USA*, **62**, 1100–1107.
Galas,D. and Schmitz,A. (1978) *Nucleic Acids Res.*, **5**, 3157–3170.
Greenblatt,J. and Schleif,R. (1971) *Nature*, **233**, 166–170.
Hahn,S. and Schleif,R. (1983) *J. Bacteriol.*, **155**, 593–600.
Hamilton,E. and Lee,L. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 1749–1753.
Harvey,S. and Cheung,H. (1982) *Cell Muscle Motil.*, **2**, 279–302.
Hendrickson,W. and Schleif,R. (1984) *J. Mol. Biol.*, **178**, 611–628.
Hendrickson,W. and Schleif,R. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3129–3133.
Hochschild,A. and Ptashne,M. (1986) *Cell*, **44**, 681–687.
Huo,L., Martin,K. and Schleif,R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 5444–5448.
Jordan,S.R. and Pabo,C.O. (1988) *Science*, **242**, 893–899.
Lee,D.-H. and Schleif,R. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 476–480.
Lee,N. and Menon,K. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 3708–3712.
Lee,N., Francklyn,C. and Hamilton,E., (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8814–8818.
Lobell,R. and Schleif,R. (1990) *Science*, **250**, 528–532.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Marmorstein,R., Carey,M., Ptashne,M. and Harrison,S. (1992) *Nature*, **356**, 408–414.
Martin,K. (1987) Ph.D. thesis, Brandeis University, Waltham, MA.
Martin,K., Huo,L. and Schleif,R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3654–3658.
Menon,K. and Lee,N. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 3708–3712.
Ogden,S., Haggerty,D., Stoner,C., Kolodrubetz,D. and Schleif,R. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 3346–3350.
Pabo,C., Sauer,R., Sturtevant,J. and Ptashne,M. (1979) *Biochemistry*, **76**, 1608–1612.
Schleif,R. (1987) In Neidhardt,F., Ingraham,J., Low,K.B., Magasanik,B., Schaecter,M. and Umbarger,H.E. (eds) *Escherichia coli and Salmonella typhimurium*. Am. Soc. Microbiol., Washington DC, Vol 2, pp. 1473–1481.
Schleif,R. (1992) *Annu. Rev. Biochem.*, **61**, 199–223.
Schleif,R. and Wensink,P. (1981) *Practical Methods in Molecular Biology*. Springer-Verlag, New York.
Smith,D. and Johnson,A. (1992) *Cell*, **68**, 133–142.
Steffen,D. and Schleif,R. (1977) *Mol. Gen. Genet.*, **157**, 341–344.
Sosnick,T., Benjamin,D., Novotny,J., Seeger,P. and Trewhella,J. (1992) *Biochemistry*, **31**, 1779–1786.
Thompson,C.C., Brown,T.A. and McKnight,S.L. (1991) *Science*, **253**, 762–768.
Tobin,J. and Schleif,R. (1990) *J. Mol. Biol.*, **211**, 75–89.
Weiss,M.A., Karplus,M., Patel,D.J. and Sauer,R.T. (1983) *J. Biomol. Struct. Dyn.*, **1**, 151–157.
Wilcox,G. and Meuris,P. (1976) *Mol. Gen. Genet.*, **145**, 97–100.
Wu,H.M. and Crothers,D.M. (1984) *Nature*, **308**, 509–513.

Received on August 7, 1992; revised on October 7, 1992