THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 7, Issue of February 16, pp. 4886–4888, 2001

Printed in U.S.A.

The C-terminal End of AraC Tightly Binds to the Rest of Its Domain*

Received for publication, August 30, 2000, and in revised form, November 13, 2000
Published, JBC Papers in Press, December 4, 2000, DOI 10.1074/jbc.M007956200

Tara L. Harmer and Robert Schleif‡

From the Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218

Genes were synthesized to express two DNA binding domains of AraC connected by short linkers. The abilities of the resulting proteins to bind to DNA containing AraC half-sites separated by the usual four bases as well as an additional two or three helical turns of the DNA were measured. The inability of some of the protein constructs to bind to widely separated half-sites indicates that the C-terminal 14 amino acids of AraC are firmly bound to the rest of the DNA binding domain.

More than 100 different proteins currently can be found in sequence data bases to possess significant sequence identity and a similarity to the DNA binding domain of AraC, the regulator of the arabinose operon in *Escherichia coli*. The pairwise identity between these domains often is below the 25% threshold for assurance that a pair of proteins possesses the same tertiary structure (1, 2). Nonetheless, the fact that the family of homologs is so large, and the fact that many whose functions have been determined are transcriptional regulators and that similarity extends across 120 amino acids all greatly raise the likelihood that the family members have the same structure. The structures of two members, MarA (3) and the DNA binding domain of Rob (4), with a 40% identity between the two and 16 and 20% identity, respectively, to the DNA binding domain of AraC have been determined and do have the same structure.

AraC has a poor similarity with MarA and Rob over its final 14 amino acids residues 282–292 (Fig. 1). Is this region a structureless tail extending into the solvent, or is it firmly bound to the protein as is the same sequence region of MarA and Rob? This information is important in the construction of variants of AraC in which arms must be added to the protein to enable it to bind to other proteins. The engineering of such arm domain interactions is appealing in the construction of alternative regulation schemes (5).

As of yet, no structure has been reported for the DNA binding domain of AraC. Thus, structure information about the DNA binding domain of AraC and particularly whether the C-terminal amino acids of the protein are free in the solvent or bound to the rest of the domain must be obtained by indirect means. By connecting two DNA binding domains of AraC with arms connecting to different places on the DNA binding domain, we have been able to show that the final 14 amino acids are firmly bound to the rest of the domain.

* This work supported by National Institutes of Health Grant GM18277. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Dept. of Biology, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218. Tel.: 410-516-5206; Fax: 410-516-5213; E-mail: bob@gene.bio.jhu.edu.

EXPERIMENTAL PROCEDURES

Genes coding for the double DNA-binding proteins and DNA used in the DNA migration retardation assays were synthesized by polymerase chain reaction using oligonucleotides approximately 60 nucleotides long, purchased from Integrated DNA Technologies, synthesized at the 200 nM scale, purified by polyacrylamide gel electrophoresis, and dissolved in dH_2O at 1 mg/ml.

The sequences of the I_1I_2, I_1-*21*-I_1, and I_1-*32*-I_1 DNAs used in the migration retardation assay were TTTGCTAGCCCTAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACATATCGCGTTTACT, TTTGCTAGCCCTAGCATTTTTATCCATAGCTACTGGTACCGTCTCATGGAAATTAGCATTTTTATCCATATCGCGTTTACT, and TTTGCTAGCCCTAGCATTTTTATCCATAGCCTAGCGTCGTTGACTGGTACGGTCTCACAG AGATTAGCATTTTTATCCATATCGCGTTTACT, where the I_1 or I_2 half-sites are underlined.

The double DNA-binding proteins were constructed using the two-step polymerase chain reaction protocol as described previously (6). All genes described here were cloned into pSE380, a 4.4-kilobase plasmid containing the promoter p_{trc}, a *ColE*1 origin of replication, *lacO* operator, $LacI^q$ repressor gene, and ampicillin resistance (Invitrogen, San Diego, CA). Plasmid constructs are summarized in Table I.

The first DNA binding domain of each double domain protein consists of AraC amino acids 169–283 at least, and the second DNA binding domain consists of AraC amino acids of at least 175–292. For the linkers connecting the DNA binding domains, we chose amino acids from the linker region of the yeast mating-type repressor protein, α2 (7). The linker provides the yeast repressor protein with sufficient flexibility that it can bind to a variety of spacings and orientations of its DNA half-sites (8). The linker was flanked on either side by two amino acids dictated by the restriction endonuclease sites *Eco*RI and *Bam*HI. The proteins were expressed in strain SH321 (F^-, $araC^-leu1022$, *lac74, galK*$^-$, str^r) (9).

Arabinose isomerase was assayed as described (10). Cells were grown in liquid minimal salts media with 0.4% glycerol and 0.4% casamino acids to an apparent optical density at $A_{550\ nm}$ of between 0.300 and 0.600 in the presence and absence of 0.2% arabinose as noted.

DNA migration retardation assays were performed as described (11). Proteins were overexpressed in a 5-ml culture, which had been inoculated with 16 μl of stationary phase SH321 cells and then grown in the absence followed by the presence of isopropyl-1-thio-β-D-galactopyranoside for 90 min. Cells were pelleted by centrifugation, resuspended in 300-μl buffer containing 100 mM K_2PO_4, 50 mM KCl, 10% glycerol, 1 mM EDTA, 1 mM dithioerythritol, and 1 mM phenylmethyl sulfonylfluoride, and then lysed by sonication. Samples were centrifuged, and the supernatant was added to 175 μl of 100% glycerol. Western transfer experiments coupled with stained SDS gels show that the various linker proteins constitute about one-twentieth of 1% of the protein in the lysates. Radiolabeled DNA template was incubated in the presence of calf thymus DNA, arabinose, and buffer containing 50 mM KCl, 25 mM Na-Hepes, pH 7.4, 2.5 mM dithioerythritol, 0.1 mg/ml bovine serum albumin, 0.1 mM K-EDTA, 5% glycerol, and 1% arabinose. 1 μl of cell lysate was added, and the reaction was incubated at 37 °C for 10 min and then loaded on a gel containing 6% acrylamide, 0.1% bisacrylamide, 0.1% ammonium persulfate, and 0.2% *N,N,N′N′*-tetramethylethylenediamine. The gel was run in a recirculated chilled buffer containing 10 mM Tris acetate and 1 mM K-EDTA, pH 7.4, at 75 V for 1 h. Gels were dried under vacuum and exposed overnight to phosphor plates. A Molecular Dynamics PhosphorImager was used to scan and analyze the gels.

FIG. 1. **AraC DNA binding domain MarA and Rob sequence alignment.** Alignment performed using ClustalW 1.8 with a gap extension penalty of 3 is shown (14).

TABLE I
Plasmids constructed

The connector shows the final amino acid of the first AraC DNA binding domain, the connector sequence including any natural AraC linker, and the first amino acid of the second AraC DNA binding domain.

Plasmid	Connector
pTH10	^{292}EFAGSINESLHPP^{177}
pTH11	^{292}EFAGSPP^{177}
pTH13	^{283}EFAGSPP^{177}
pTH15	^{286}LARQVVRIQCGNSRLINGSINESLHPP^{177}
pTH16	^{286}LAVKLSEFAVVRLINGSINESLHPP^{177}

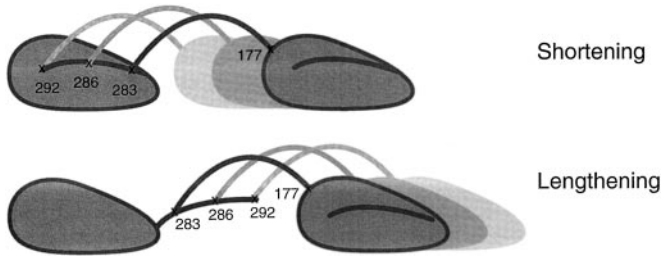

FIG. 2. **Two AraC DNA binding domain protein linkers attached to different points on the first DNA binding domain.** If the C-terminal stretch is tightly bound as shown in the *top diagram*, the effect is shortening, but if the C-terminal stretch is not bound, the effect is lengthening as shown in the *bottom diagram*.

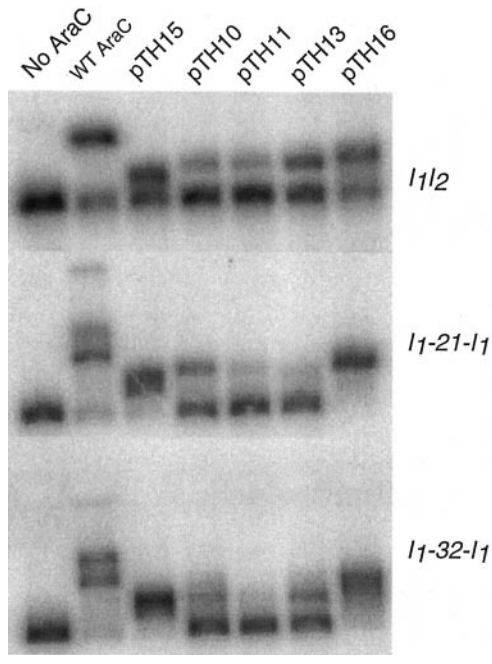

FIG. 3. **Binding of AraC and AraC derivative DNA binding domain proteins to three DNA templates, I_1I_2 , I_1-*21*-I_1, and I_1-*32*-I_1.**

RESULTS

The C terminus of MarA consists of amino acids 111–124, corresponding to AraC amino acids 278–291. This C-terminal stretch lies across the side of the protein opposite the DNA binding site. If the corresponding C-terminal stretch in AraC is attached tightly to the core, shifting the attachment point of the linker to the first DNA binding domain toward the C-terminal end of the protein should effectively shorten the maximum distance that the two DNA binding domains can be spaced when binding to direct repeat half-sites as illustrated in Fig. 2. If instead, the C-terminal stretch is loosely attached as also shown in Fig. 2, moving the attachment point toward the C terminus of the protein effectively lengthens the linker and increases the maximum distance that the two DNA binding domains can be separated when binding in a direct repeat orientation.

To test whether the C-terminal stretch of AraC is held firmly against the body of the DNA binding domain or whether it is loosely attached, we performed a series of DNA binding assays with the double AraC DNA binding domain proteins and three DNA templates, wild type *ara* I_1I_2, I_1-*21*-I_1, which contains two helical turns between the half-sites, and I_1-*32*-I_1, which contains three helical turns. If the C-terminal stretch is held tightly to the body, shifting the linker attachment point to the first DNA binding domain from AraC amino acids 263 to 292 will result in a shortening of the maximum distance of separation between the centers of the structured cores of the two DNA binding domains and decrease the binding ability to widely spaced DNA binding sites. However, if the association is weak, shifting the linker attachment point will result in a lengthening of the maximum separation among the centers of the structured cores of the DNA binding domains. This will allow the protein to bind to DNA with more widely separated half-sites. Fig. 3 and Table II show that pTH11 binds not as well as pTH13 does with increasing distance between the binding sites. This result indicates that the C-terminal stretch, indeed, is held firmly to the body of the protein.

DISCUSSION

In this work, we have shown that the C terminus of AraC residues 278–292 binds tightly to the body of the AraC DNA binding domain. This was determined by changing the attachment points of the linker to the first DNA binding domain of double AraC DNA binding domain proteins, assaying the binding abilities of the resulting proteins to DNA templates with AraC half-sites separated by the wild-type spacing, and assaying the binding abilities of the resulting proteins to DNA tem-

TABLE II
Summary of results

Linker length is the number of amino acids between attachment points to the two AraC DNA binding domains. Attachment point indicates the AraC amino acid position where the linker attaches to the first DNA binding domain. The maximum reach is the maximum distance between the centers of the structured cores of the two DNA binding domains when binding to half-sites in a direct repeat orientation when the C-terminal 14 amino acids are strongly bound to the body of the DNA binding domain. Induction at $I_1I_2p_{BAD}$ was measured by arabinose isomerase assay, and the results are reported as the percent of wild-type AraC activity. Binding to template DNA was determined by *in vitro* DNA migration retardation assays, and the number reported is the fraction of DNA found in the protein-DNA complex.

Protein	Linker length	Attachment point	Maximum reach	Induction $I_1I_2p_{BAD}$	Binding to template DNA		
					I_1I_2	I_1-*21*-I_1	I_1-*32*-I_1
	aa[a]		*Å*	%			
pTH11	7	292	60	300	0.11	0.08	0.03
pTH13	7	283	90	510	0.35	0.14	0.35
pTH10	13	292	90	370	0.15	0.34	0.22
pTH16	25	286	160	190	0.52	0.99	0.96
pTH15	27	286	170	88	0.49	0.96	1.0

[a] *aa*, amino acid.

plates with the half-sites separated by two and three additional helical turns of the DNA. The binding of the C-terminal stretch to the body of AraC is strong enough that this interaction remains intact when attached to a second AraC DNA binding domain via a flexible linker and when this entire protein is bound to DNA. The binding energies required to bind a double AraC DNA binding domain protein to I_1 half-sites separated by three helical turns are insufficient to strip the C-terminal stretch from the body of the DNA binding domain.

This work also provides data against a proposal for the mode of dimerization of AraC in the absence of arabinose and the mechanism of response to arabinose made on the basis of the crystal packing arrangements of the dimerization domain of AraC in the presence and absence of arabinose (12). AraC protein was proposed to loop between distal DNA half-sites and repress the *araBAD* operon in the absence of arabinose when dimerized by a face to face interaction that allowed the DNA binding domains to be separated by 140 Å or greater and to cease looping and activate transcription when the separation distance was decreased by a shift in the dimerization interface. The proteins encoded by pTH15 and pTH16 allow greater than 140-Å separations, and as shown in Table II, strongly induce the *ara* p_{BAD} promoter. Other data also show that the alternative interface mechanism is inapplicable to AraC (13).

REFERENCES

1. Sander, C., and Schneider, R. (1991) *Proteins* **9,** 56–68
2. Holm, L., Ouzounis, C., Sander, C., Tuparev, G., and Vriend, G. (1992) *Protein Sci.* **1,** 1691–1698
3. Rhee, S., Martin, R. G., Rosner, J. L., and Davies, D. R. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 10413–10418
4. Kwon, H. J., Bennik, M. H., Demple, B., and Ellenberger, T. (2000) *Nat. Struct. Biol.* **7,** 424–430
5. Schleif, R. (1999) *Proteins* **34,** 1–3
6. Prytulla, S., Dyson, J. H., and Wright, P. E. (1996) *FEBS Lett.* **399,** 283–289
7. Astell, C., Ahlstrom-Jonasson, R., and Smith, M. (1981) *Cell* **27,** 15–23
8. Sauer, R. T., Smith, D. L., and Johnson, A. D. (1988) *Genes Dev.* **2,** 807–816
9. Hahn, D., Dunn, T., and Schleif, R. F. (1984) *J. Mol. Biol.* **180,** 61–72
10. Schleif, R. F., and Wensink, P. (1981) *Practical Methods in Molecular Biology*, Springer-Verlag New York Inc., New York
11. Hendrickson, W., and Schleif, R. (1984) *J. Mol. Biol.* **174,** 611–628
12. Soisson, S. M., MacDougall-Shackleton, B., Schleif, R., and Wolberger, C. (1997) *Science* **276,** 421–425
13. Saviola, B., Seabold, R., and Schleif, R. F. (1998) *J. Mol. Biol.* **278,** 539–548
14. Thompson, J. D., Higgins D. G, and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680