Biochemistry

Article

Cite This: *Biochemistry* 2019, 58, 2867−2874

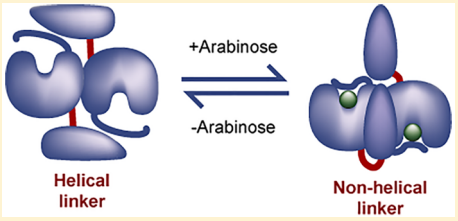
pubs.acs.org/biochemistry

# Helical Behavior of the Interdomain Linker of the *Escherichia coli* AraC Protein

Matthew J. Brown and Robert Schleif*

Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218, United States

**ABSTRACT:** In *Escherichia coli*, the dimeric AraC protein actively represses transcription from the L-arabinose *araBAD* operon in the absence of arabinose but induces transcription in its presence. Here we provide evidence that, in shifting from the repressing to the inducing state, the behavior of the interdomain linker shifts from that of an $\alpha$ helix to that of a more flexible form. *In vivo* and *in vitro* experiments show that AraC with a linker sequence that favors helix formation is shifted toward the repressing state in the absence and presence of arabinose. Conversely, AraC containing a linker sequence that is unfavorable for helix formation is shifted toward the inducing state. Experiments in which the presumed helical linker is shortened or lengthened, protein helical twist experiments, are also consistent with a helix transition mechanism. Previous experiments have shown that, upon the binding of arabinose, the apparent rigidity with which the DNA binding domains of AraC are held in space decreases. Thus, arabinose likely controls the stability or rigidity of the interdomain linker. Circular dichroism experiments with peptides show that the helicity of the linker sequence can be controlled by the helicity of residues preceding the linker, providing a plausible mechanism for arabinose to control the repressing−inducing state of AraC protein.






Interdomain linkers have been studied in many proteins and have been implicated in numerous roles, including tethering, and signal propagation.[1−4] The AraC protein presents a case in which the linker structure is directly involved with a large-scale structural change that dramatically affects the DNA binding activity of the protein.

In *Escherichia coli*, expression of the genes required for the catabolism of the sugar L-arabinose, *araBAD*, is controlled by the *araC* gene product.[5] AraC functions as a homodimer, with each subunit composed of two major domains[6] (Figure 1). The first is a 166-residue dimerization domain, the first 20 residues of which comprise an N-terminal arm. The C-terminal domain is a 117-residue DNA binding domain. This DBD is connected to the dimerization domain by an eight-residue interdomain linker.[7,8]

In the absence of arabinose, the AraC dimer contacts the $I_1$ DNA half-site and the $O_2$ half-site lying 210 bp upstream from $I_1$[9] (Figure 1). AraC binding to $I_1$ and $O_2$ loops the DNA. This interferes with the binding of RNA polymerase to the nearby *ara* $p_{BAD}$ promoter,[10] thereby repressing its activity. A third binding site, $I_2$, lies adjacent to $I_1$ in a direct repeat orientation, partially overlapping the $p_{BAD}$ core promoter.[9,11] It is unoccupied in the minus-arabinose, repressive, state of the operon,[12] but when arabinose is added, the DNA binding domain previously at the $O_2$ half-site relocates to the $I_2$ half-site, eliminating the DNA looping.[13] Following its repositioning on the DNA, AraC stimulates *araBAD* expression from the $p_{BAD}$ promoter[13,14] by assisting the DNA binding and isomerization of RNA polymerase to form an open complex.[15] Therefore, DNA looping hinders the activity of $p_{BAD}$ in two ways: first, by interfering with access of the polymerase to the promoter, and second, by holding and preventing a DNA binding domain of AraC from binding to $I_2$.

Previous experiments have shown that reversing the direction of the asymmetric $O_2$ half-site while maintaining its position on the same face of the DNA reduces or eliminates the capacity of AraC to loop to $O_2$ from $I_1$.[16] This, plus *in vitro* DNA binding experiments using direct repeat and inverted repeats of the half-sites,[11] shows that in the absence of arabinose, the relative positions and/or orientations of the DNA binding domains of AraC are constrained in a way that favors DNA looping between $O_2$ and $I_1$ as well as hindering their binding to direct repeat half-sites. Finally, *in vitro* DNA binding experiments using DNA containing two AraC half-sites separated by single-stranded DNA lead to a similar conclusion.[17] That is, on the binding of arabinose, the protein's apparent rigidity is reduced, thus changing the lower-energy state of the system from the repressive, DNA looping, state to the inducing state with AraC bound to the $I_1$−$I_2$ direct repeat DNA half-sites. Undoubtedly, in solution, multiple conformations of AraC are all present, predominately represented by inducing and repressing states that are in equilibrium. Arabinose changes the relative amounts of these species, but generally, both will be present. For the sake of convenience, however, henceforth we shall speak of the inducing or repressing state when that species is expected to be the predominant one.

The facts mentioned above raise two fundamental mechanistic questions. How is the rigidity that maintains the

**Received:** March 18, 2019
**Revised:** May 29, 2019
**Published:** June 14, 2019




DOI: 10.1021/acs.biochem.9b00234
*Biochemistry* 2019, 58, 2867−2874

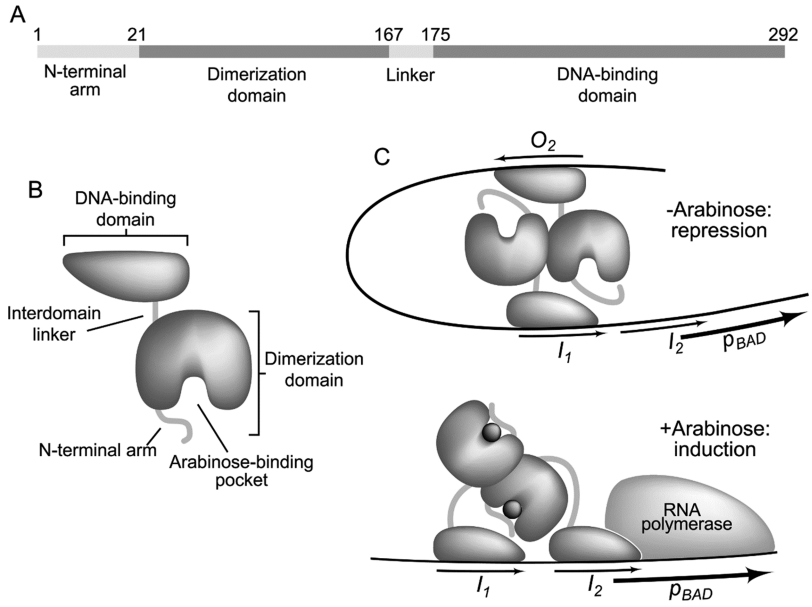



**Figure 1.** (A) Linear layout of the AraC protein domains. (B) Domain structure of AraC. (C) DNA binding sites in the *ara* regulatory region made by AraC in the absence and presence of arabinose. The $p_{BAD}$ is the RNA polymerase binding site.

orientations of the DNA binding domains in the absence of arabinose generated, and then, how is the rigidity relaxed when arabinose is added? The 20 N-terminal amino acids of AraC are an attractive candidate for controlling the rigidity. The arms' locations in the protein's three-dimensional structure are compatible with their binding to the DBDs in the absence of arabinose, and this interaction could then hold the DNA binding domains in positions suitable for DNA looping.[7] Furthermore, in the absence of arabinose, the arms are available to bind to the DNA binding domains,[18] but in the presence of arabinose, the arms are draped over the dimerization domain, with the side chains of F15 reaching into the domains and making direct contact with the bound arabinose.[7]

Although arm−DNA binding domain interactions in the absence of arabinose appear to be an attractive regulatory mechanism, multiple experiments to detect such contacts have failed to do so (to be discussed below). Experimental attention has therefore shifted from mechanisms involving direct arm−DBD interactions to less direct mechanisms for holding the DNA binding domains. *In vitro* experiments have recently suggested that the interdomain linker plays a direct role in the rigidity shift of AraC.[19] These experiments utilized truncated AraC consisting of the dimerization domain, including the N-terminal arm and the eight-residue interdomain linker, but lacking the DNA binding domain. Fluorescence anisotropy measurements using such a truncated AraC with a fluorescent label at the C-terminal end of the linker showed that the label's rotational motion is controlled by arabinose.[19] In the presence of arabinose, its level of motion is increased, as though the linker's rigidity is reduced by the presence of arabinose. Additionally, a linker's apparent flexibility is controlled by only the N-terminal arm that is adjacent to the linker, and not by the other arm,[19] implying that it is a relatively direct interaction between the arm and linker that controls flexibility rather than a global structural change.

In the work to be described below, we address the genetic hint contained in earlier work suggesting that, in the repressing state, the interdomain linker is largely in an $\alpha$-helical form, and the presence of arabinose shifts the linker toward a nonhelical state.[20] This mechanism is well suited to the rigidity shift mentioned above. Here we show that increasing the helical propensity of the linker strengthens the tendency of AraC to be in the repressing state as shown by a reduction in the inducibility by AraC *in vivo* and by a reduction in the affinity for binding to adjacent direct repeat half-sites ($I_1$−$I_2$) *in vitro*. Conversely, decreasing the helical propensity of the linker weakens the tendency to be in the repressing state as shown by increases in *in vivo* inducibility and uninduced basal levels as well as increasing the protein's *in vitro* DNA binding affinity to direct repeat half-sites. To provide additional data relevant to a helix−coil mechanism, we have performed a protein helical twist experiment with length changes to the linker of −2, −1, 0, +1, and +2 residues. In these experiments, if the linker were helical, the addition or removal of a single residue from each interdomain linker would rotate the two DNA binding domains by ∼180° with respect to each other and would be expected to impair repression. Similarly, the addition or removal of two residues would bring the two DNA binding domains back close to their original relative orientation and could partially restore normal repression. Such results were found. We have also obtained circular dichroism spectra showing that, although a peptide reproducing the linker sequence is largely unstructured in solution, it becomes helical when a longer helical peptide sequence is added to it. In AraC, the linker immediately follows a helix in the protein's dimerization domain, and hence, the dimerizing helix could nucleate helix formation of the linker in the protein. Furthermore, this suggests that control of the linker's structural state may be modulated by that of the immediately preceding residues.

## MATERIALS AND METHODS

**Plasmids and Strains.** Except where noted, mutants of AraC were generated using a Stratagene QuikChange protocol for site-directed mutagenesis (Stratagene, La Jolla, CA) from oligonucleotide primers and verified by sequencing the products. The AraC variants used in this paper additionally



DOI: 10.1021/acs.biochem.9b00234
*Biochemistry* 2019, 58, 2867−2874

contained the Y31V mutation,[18] which reduces the level of aggregation and facilitates *in vitro* experiments.

For the *in vivo* activity measurements of the 6A, GS linker variants, the AraC protein was expressed from the pWR03 AraC expression vector based on pSE380.[21] The linker spacing mutations were made in the *araC pBAD*-GFP vector, derived from the pBR322 plasmid and formerly available from Clontech. This places expression of GFP under control of the $p_{BAD}$ promoter and associated full regulatory region. The *pBAD*-GFP vector was modified to contain the two DNA replication origins of the pSB2K3 vector described at http://parts.igem.org/Plasmid_backbones/Operation#Inducible_copy_number_operation_plasmid_backbones. In strains containing the *lac* $i^Q$ mutation in the absence of isopropyl β-D-1-thiogalactopyranoside (IPTG), the vector, pCCNC, is present at a copy number of <10 and in the presence of 1 mM IPTG at a copy number of >100. The GFP levels resulting from uninduced *ara* $p_{BAD}$ cannot easily be measured due to other fluorescence sources within cells. Therefore, for measurement of activity from AraC variants carried on this plasmid, it was transformed into a strain without AraC but containing *ara* $p_{BAD}$ and the *araBAD* genes, thereby allowing measurement of arabinose isomerase for quantitation of the $p_{BAD}$ basal levels. The strain also contained the $i^Q$ mutation that is required to hold the plasmid copy number at a low value. This strain, RS321, is derived from SH321[22] by P1 transduction transferring Tn10 (laci$^Q$ *Cm*$^r$ at the *lac* locus) from Gottesman strain ASP7020 (NM514).[23]

**DNA Used in Binding Measurements.** The $I_1$−$I_1$ oligomers used in the DNA binding assays were 5′-/Cy5/gccaTAGCATTTTTATCCATAagatTAGCATTTTTATCCATAcctc-3′ as well as the unlabeled reverse complement. Underlined regions indicate $I_1$ binding sites. A specific competitor was the same sequence but lacked the Cy5 label. Labeled DNA was annealed in 25 mM Tris (pH 8), 1 mM EDTA, 2.5 mM MgCl, and 100 mM KCl at a 1:1.1 molar ratio of labeled to unlabeled strands to ensure that no fluorescent label remained single-stranded. An unlabeled competitor was annealed at an equimolar ratio of both strands. Annealing was performed by incubation at 95 °C for 5 min before the temperature was gradually decreased to 20 °C over a period of 30 min.

**Protein Used in DNA Binding Measurements.** For the *in vitro* DNA binding measurements, AraC was overproduced using the pET-24 expression vectors (Novagen) with AraC cloned between the *Nco*I and *Sac*I restriction sites and isolated by ion exchange chromatography sequentially on heparin, HiTrap-Q, and again heparin columns as described by Rodgers and Schleif.[24]

**Arabinose Isomerase Assay.** Arabinose isomerase assays were performed as described by Schleif and Wensink.[25] Cells, containing plasmid SH321 (*F*$^-$ Δ*araC-leu1022* Δ*lac74 galK*$^-$ *Str*$^r$ *thi*$^-$1$^4$) or SH322 (*F*$^-$ Δ*lac74 galK*$^-$ *Str*$^r$ *thi*$^-$1$^4$),[22] were grown in M10 minimal salts medium[25] with 0.4% glycerol or 0.4% arabinose, 10 μg/mL thiamine, 0.2% casamino acids, and 100 μg/mL ampicillin to an apparent $OD_{650}$ of 0.6−0.8 (as measured by a Perkin/Elmer Lambda 25 UV/vis spectrophotometer) before harvesting and concentration for an assay of isomerase. Cells containing pCCNC were grown to stationary phase in YT medium containing 40 μg/mL kanamycin for measurement of uninduced levels of arabinose isomerase.

**DNA Dissociation and Electrophoretic Mobility Shift Assay.** The binding−dissociation experiments were performed at 37 °C in 10 mM Tris (pH 7.4), 1 mM K-EDTA, 1 mM DTT, 5% glycerol, and 0.1 mg/mL BSA containing either 100 mM KCl for the minus-arabinose measurements or 300 mM KCl and 5 mM arabinose for the plus-arabinose experiments. These concentrations were chosen to provide an optimum time scale for experimentation, allowing for sufficient resolution while removing the complications introduced by longer time courses. DNA and protein were equilibrated together using 50 nM labeled $I_1$−$I_1$ DNA and 150 nM purified AraC protein in 10 μL volumes for 5 min prior to the addition of a 30-fold excess of unlabeled $I_1$−$I_1$ DNA and 0.5 μg of sheared calf thymus DNA in 10 μL. Samples were then loaded and immediately run on 6% acrylamide, 0.1% methylene bisacrylamide gels containing 10 mM Tris-acetate (pH 7.4) and 1 mM K-EDTA at 8 V/cm for 30 min. Control experiments showed that DNA entered gels within 30 s of loading. Gels were imaged on a GE Typhoon 9410 fluorescence scanner using a 633 nm laser for excitation and a 670 nm emission filter. The resultant images were analyzed using GE ImageQuant TL 7.0 software.

The data from both the plus- and minus-arabinose experiments show the presence of a small amount (<10%) of a rapidly dissociating bound species. We do not know the origin of this species and have ignored it in our analyses. This rapidly dissociating component has appeared in multiple past experiments.

**Circular Dichroism (CD) Measurements and Sample Preparation.** N-Terminally acetylated peptides for CD measurements were obtained from GenScript USA Inc. as a crude preparation and further purified by reverse phase liquid chromatography using a PepRPC 15 μm column, eluting with acetonitrile, and then dried under vacuum by centrifugal evaporation. CD spectra were recorded on an Aviv 2.86 spectropolarimeter. Samples were prepared in 10 mM phosphate buffer (pH 7) and 100 mM NaCl. Spectra were recorded at 2.0 °C at 1 nm intervals from 250 to 185 nm, averaging over 5 s. The Dynode exceeded 350 V at wavelengths of <190 nm, causing considerable uncertainty in this region. As such, these measurements were discounted during analysis. To obtain the reported values, the CD signal from buffer alone was subtracted from all spectra, after which values were converted into mean residue ellipticity units.

## RESULTS

***In Vivo* Properties of Helix-Stabilizing and -Destabilizing Mutants.** To increase the helical preference of the eight-residue linker [wild type (WT) sequence INESLHPP], we replaced its first six residues with alanine, the residue with the highest individual helical propensity,[26] giving an AAAAAAPP linker sequence (6A). Conversely, we also constructed a glycine-serine linker, GGGGSGPP (GS), to decrease the helical propensity of the sequence and bias its structure toward a random coil.[26]

Table 1 shows the *in vivo* properties, that is, the repressing and inducing activities of wild type AraC and AraC containing the 6A and GS linkers. In the case of AraC with 6A, the helix-favoring linker, uninduced levels remained comparable to those seen in WT AraC, while induced levels were reduced by a factor of 2. Conversely, the GS, random coil-favoring, linker mutant showed both a greatly elevated uninduced level, that is a loss of repression, and an elevated induced level. These results are consistent with the idea that at least part of the



DOI: 10.1021/acs.biochem.9b00234
*Biochemistry* 2019, 58, 2867−2874

Table 1. *In Vivo* Activities of AraC Linker Mutants on $arap_{BAD}$[a] in Units per Cell[25]

| | minus arabinose | plus arabinose |
|---|---|---|
| WT | 4.6 ± 1.7 | 960 ± 190 |
| 6A | 5.4 ± 1.6 | 510 ± 110 |
| GS | 130 ± 5 | 3100 ± 230 |

[a]Values are the average and standard deviation of five independent measurements of arabinose isomerase levels in AraC$^-$ cells expressing the indicated form of *araC* from a plasmid.

linker is helical in the repressing state and is not helical in the inducing state.

**Excluding a Potential Artifact.** The weakened ability of AraC containing the 6A linker to induce could result from one of three possibilities: (1) as in our hypothesis, a shift toward the repressing conformation in the arabinose-controlled equilibrium between the repressing and inducing conformations, (2) reduced total levels of active protein, and (3) a weakened ability to activate transcription once bound at *araI*.

It has previously been shown that the repressing state of AraC is dominant with respect to the inducing state,[3] which is an expected consequence of the DNA looping mechanism of the system. Therefore, if the 6A AraC has been pushed toward the repressing state, its hypo-inducible phenotype will be dominant over the fully inducible wild type. Such dominance would not be seen for the other two possible reasons for hypo-inducibility. Table 2 shows that hypo-inducibility is dominant, and hence, the behavior of the linker variants is consistent with the helix−coil hypothesis.

Table 2. Hypo-Inducibility of the 6A Linker That Is Dominant with Respect to Wild Type AraC[a]

| AraC linker | minus arabinose | plus arabinose |
|---|---|---|
| WT | 4.3 ± 0.9 | 1100 ± 235 |
| 6A | 6.0 ± 2.2 | 625 ± 80 |

[a]Units of arabinose isomerase per cell from three independent measurements in heterozygous cells containing wild type *araCBAD* genes on the chromosome and the indicated mutant AraC on the plasmid.

***In Vitro* Properties of the Helix-Stabilizing and -Destabilizing Variants.** As mentioned above, the binding affinity of AraC *in vitro* for adjacent direct as well as inverted half-sites has been studied.[11] The results indicate that the AraC DNA binding domains are significantly constrained in the absence of arabinose to positions and orientations that are unfavorable for binding to adjacent, DNA half-sites but that in the presence of arabinose they are less constrained. Thus, in the absence of arabinose, the net free energy available for binding to adjacent DNA half-sites is reduced by the amount of energy required to overcome the constraints. In the presence of arabinose, less energy is required to overcome the constraints, and AraC can bind the adjacent direct repeat DNA more tightly. Consequently, if the interdomain linker in wild type AraC in the repressing state were helical and breaks for the protein to bind to the DNA sites required for induction, then an interdomain linker sequence that increases the stability of the helix will weaken the binding of AraC to adjacent DNA half-sites and a linker sequence that decreases the stability of the helix will do the opposite. The amount of the changes in affinity cannot be predicted, nor can they be directly related to the sizes of the changes in repression or induction; thus, what is relevant in the results is whether the affinities increased or decreased.

For AraC and many other DNA binding proteins whose forward rate of DNA binding is constant and close to the diffusion limit,[27−29] relative equilibrium binding constants can be determined by comparing their more easily measured dissociation rates. We used a direct repeat of the $I_1$ half-site to further facilitate measurements as AraC dissociates more slowly from this than from the natural $I$ site ($I_1$−$I_2$).

Figure 2 shows a representative gel from which dissociation kinetics can be determined, and Figure 3 shows the

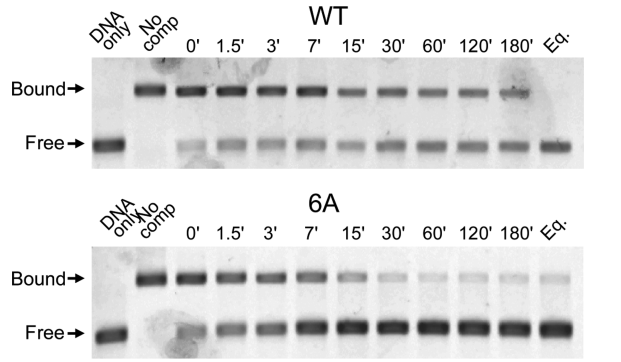



**Figure 2.** DNA electrophoretic mobility shift measuring the dissociation kinetics of AraC protein from the $I_1$−$I_1$ site in the presence of arabinose and 300 mM KCl. The first lane contains only labeled $I_1$−$I_1$ DNA. The second lane contains labeled $I_1$−$I_1$ DNA and protein to show that DNA is completely bound. For lanes 3−11, samples were incubated with unlabeled $I_1$−$I_1$ for the indicated time before being loaded into the gel. For the sample run in the final lane, competitor DNA was added before the protein to verify that, in the other samples, sufficient competitor was present to ensure that any protein that dissociated from labeled DNA would almost surely reassociate with the competitor and not the labeled DNA.

dissociation kinetics of wild type and the 6A and GS mutant linker proteins in the absence and presence of arabinose. We performed the plus- and minus-arabinose measurements under different salt conditions, as the dissociation rate was previously found to be highly salt-dependent,[30] and conditions were chosen to accommodate both convenient and reliable time scales. Nothing in the structure of AraC[7,18,31] or in multiple previous experimental studies from this lab involving variable salt concentrations suggests that the mechanisms involved might change with the salt or arabinose concentrations actually used.[20,30,32] In the absence of arabinose, the helix-favoring 6A AraC dissociates from DNA significantly faster than wild type AraC. In the presence of arabinose, the 6A AraC dissociates at approximately the same rate as wild type AraC. Conversely, the random coil favoring GS AraC dissociates more slowly from DNA than wild type AraC, in the absence and presence of arabinose. These results, also, are consistent with the idea that a linker sequence favoring a helical conformation weakens binding to adjacent DNA half-sites and are consistent with the *in vivo* results. That is, in the presence of arabinose, the equilibrium state of the linker shifts toward a more flexible, nonhelical state, which facilitates binding to adjacent half-sites.

**Protein Helical Twist Experiments.** If the interdomain linker of AraC were an $\alpha$ helix and if the two DNA binding domains point roughly in opposite directions in the repressing state, as shown in Figure 1, then the removal or the addition of



DOI: 10.1021/acs.biochem.9b00234
*Biochemistry* 2019, 58, 2867−2874

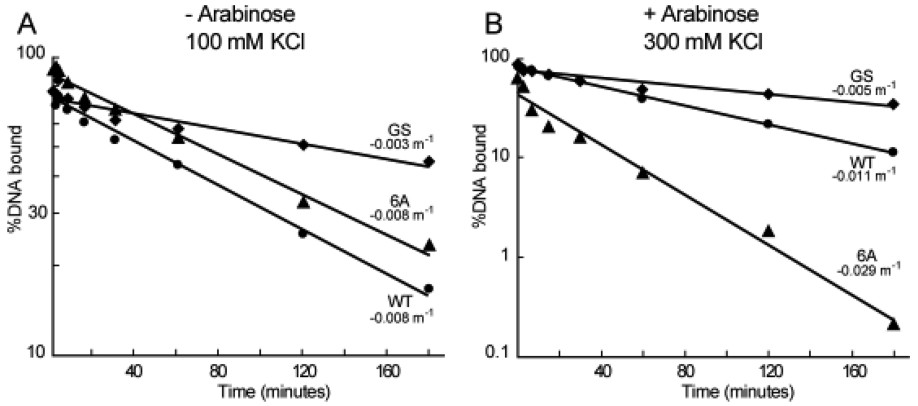



**Figure 3.** Dissociation of $I_1$–$I_1$ from AraC linker variants as a function of time. (A) Dissociation in 100 mM KCl in the absence of arabinose. (B) Dissociation in 300 mM KCl in the presence of arabinose. Plus- and minus-arabinose measurements were taken at different salt concentrations to maintain experimentally convenient dissociation rates.

a helical residue in each of the two linkers would reorient the DNA binding domains to be roughly parallel, which should interfere with formation of the repressing state. This follows from the fact that the helical pitch of an $\alpha$ helix is nearly four residues per turn. Therefore, the addition or removal of one residue from a helical linker seems likely to rotate its attached DNA binding domain ~90° about the axis of the helix. (Recall that the DNA binding domains do not appear to be free to pivot on the linkers because reversing the orientation of the $O_2$ half-site interferes with repression.[16]) A similar addition or removal of one residue from the other linker rotates the other DNA binding domain by 90°, but in the opposite direction. Consequently, the two DNA binding domains shift from pointing in roughly opposite directions, which appears to be necessary for DNA looping and repression, to pointing in nearly the same direction (Figure 4).

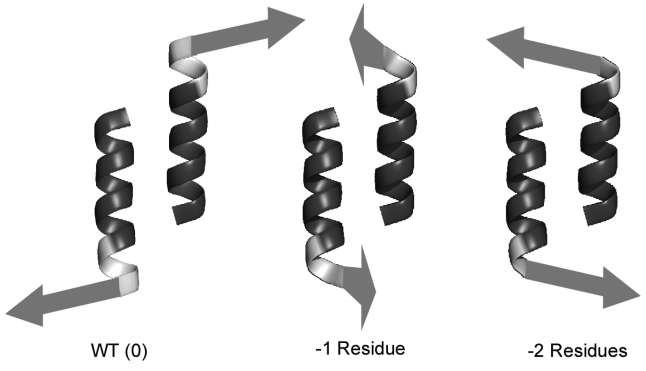



**Figure 4.** Schematic showing the rotational effects on the ends of the two helices relative to their central portions produced by the insertion or deletion of one or two residues.

This reorientation of the DNA binding domains could interfere with DNA looping and repression in a manner similar to that observed upon reversing the orientation of the $O_2$ half-site. Consequently, the basal level of *araBAD* expression should increase. The addition or removal of two helical residues from the linkers could bring the DNA binding domains back closer to their original relative orientation, and we might then expect to find the basal level decreased back toward the basal level displayed by AraC with unaltered linkers. The results listed in Table 3 show that, indeed, the protein with an insertion or deletion of one residue in each

**Table 3. Basal Levels of Arabinose Isomerase in Cells with AraC Linker Deletions and Insertions**[a]

| linker length change | experiment 1 | experiment 2 | experiment 3 | average and standard deviation |
|---|---|---|---|---|
| −2 | 11 | 8.3 | 9.8 | 9.7 ± 1.1 |
| −1 | 26 | 15 | 16 | 19 ± 4.9 |
| 0 (WT) | 16 | 14 | 11 | 13.7 ± 2.0 |
| +1 | 28 | 20 | 26 | 24.7 ± 3.4 |
| +2 | 23 | 17 | 21 | 20.3 ± 2.5 |

[a]Units of arabinose isomerase per cell from three independent experiments of RS321 [Δ*araC* (*araBAD*+)] cells containing the plasmid pCCNC *araC* with the AraC linker length changes indicated. +1 and +2 insertions were alanine residues inserted between S170 and L171. −1 and −2 deletions remove residue S170 and residues S170 and L171, respectively.

linker represses less well than AraC with linkers of wild type length or linkers containing an insertion or deletion of two residues.

**Helix Propagation into the Linker Sequence.** A sequence of six or eight residues is unlikely to be able to nucleate and form a helix on its own. If, however, it were directly preceded by a helix, as is the case in AraC, it could well be helical itself. We therefore sought to determine if a preceding helix can be extended into the linker, as would be required for a helix–coil transition mechanism. If linker helicity can be specified in this way, helicity could plausibly be regulated by breaking the preceding helix, suggesting a possible mechanism for the hypothesized transition. To test this question, we used a series of peptides and examined their helical content by CD.

We used four peptide sequences: the eight-residue linker alone, a 16-residue helical sequence alone, and this helix attached to the linker by either an alanine or a proline residue (Table 4). We expected the helix–Ala–linker sequence to allow the helix structure to continue into the linker and the helix–Pro–linker to interrupt it. Our prenucleated "seed" helix was based on the 3K(I) sequence characterized by Marqusee et al.,[33] differing only by our inclusion of an N-terminal tryptophan residue for spectrophotometric concentration measurement. The CD spectra of these peptides are shown in Figure 5. The linker sequence alone displayed a character consistent with being almost entirely unstructured. Conversely, the helix sequence showed a highly helical spectrum, as


DOI: 10.1021/acs.biochem.9b00234
*Biochemistry* 2019, 58, 2867–2874

**Table 4. Sequence Layout of Linker-Capped Peptides**

| | Helix | Linker |
|---|---|---|
| Linker | Ac- | INESLHPPW |
| Helix | Ac-WAAAAKAAAAKAAAAKA | |
| H-A-L | Ac-WAAAAKAAAAKAAAAKA**A**INESLHPPW | |
| H-P-L | Ac-WAAAAKAAAAKAAAAKA**P**INESLHPPW | |

Ac – acetyl group. Boldface indicates residue not included in either linker or helix sequence.

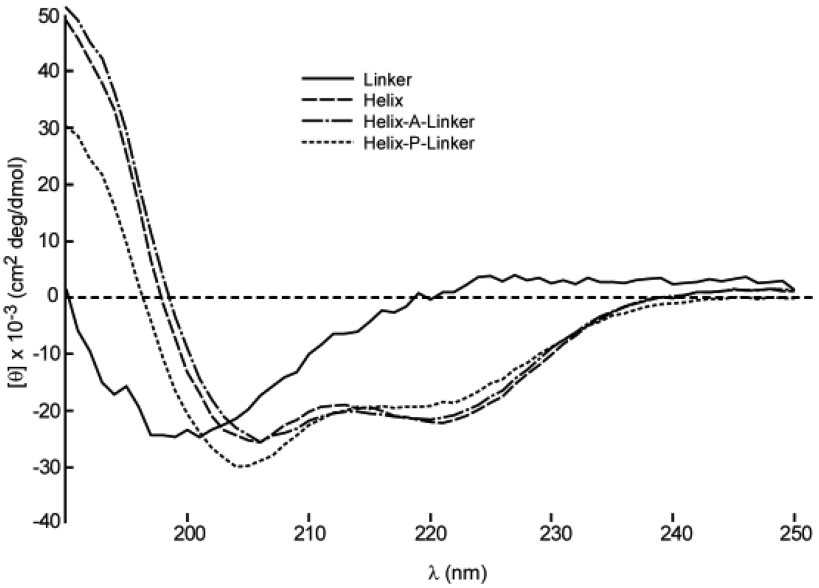



**Figure 5.** CD spectra of helix/linker peptides. All measurements were taken in 10 mM phosphate buffer (pH 7) and 0.1 M NaCl at 2.0 °C.

previously reported.[33] The helix−Ala−linker sequence showed a spectrum almost perfectly overlapping that of the helix alone, that is, no evidence of any additional random coil structure. Note particularly the two local minima around 222 and 208 nm as well as the maximum around 190 nm.[34] We therefore conclude that most of the linker sequence in this peptide is also helical. Notably, a sizable shift in the CD spectrum is seen when the linker sequence is preceded by a proline rather than an alanine, indicating that the introduction of this proline disrupts helix formation in the preceding or following helix.

## DISCUSSION

We have investigated what happens when the AraC protein shifts from its repressing and DNA looping state in the absence of arabinose to its nonlooping and inducing state in the presence of arabinose.[12,13] In the apo state, the DNA binding domains of AraC apparently are rigidly held in positions and orientations that favor binding to two well-separated DNA half-sites over binding to two adjacent DNA half-sites. The addition of arabinose reverses the binding site preference of AraC by reducing the energetic cost of binding to adjacent DNA half-sites. Thus, the protein shifts from repressing the *ara* $p_{BAD}$ promoter by DNA looping to inducing it by binding to the adjacent $I_1$ and $I_2$ half-sites.[12,13] Here we provide several types of experiments indicating that in the absence of arabinose, the interdomain linkers between the dimerization domains and DNA binding domains behave as though, at least in part, they are in an $\alpha$-helical state and that, upon binding arabinose, this behavior is reduced or lost altogether, i.e., the linker appears to have shifted from a helix to a less stable structure, perhaps a random coil. These findings, in conjunction with previous findings,[19] suggest a mechanism whereby the induction status of AraC is controlled by a conformational transition of the interdomain linker toward or away from a helical state.

While direct examination of the structure of the linker would be ideal, a number of factors hinder such studies. The structure of full length AraC determined by crystallography is not available as more than 30 years of serious efforts to crystallize the protein have not been successful. Structure determination of the 64000 molecular weight AraC protein by nuclear magnetic resonance while possible would be highly expensive and most challenging due in part to the low solubility of apo-AraC. There is a chance that direct study with cryo-EM will become possible. While we do present CD spectra using peptides reproducing the linker sequence here, using similar means to monitor the helical status of the linker in the full length protein is unfeasible because 37% of the 292-residue AraC protomer is already helical, thereby preventing detection of such comparatively small changes.



DOI: 10.1021/acs.biochem.9b00234
*Biochemistry* 2019, 58, 2867−2874

To provide evidence supporting the idea that the interdomain linker is helical in the repressing state and nonhelical in the inducing state, we both increased and decreased the stability of an α-helical state of the linker. We investigated the *in vivo* regulatory properties and *in vitro* DNA binding properties of wild type AraC and of two variants: one whose interdomain linker favors helix formation and one whose linker disfavors helix formation. Compared to wild type AraC whose linker sequence is INESLHPP, the protein containing a helix-stabilizing[26] alanine-substituted linker sequence, AAAAAAPP, shifts AraC toward its repressing state and weakens its binding to adjacent direct repeat half-sites. AraC containing a sequence disfavoring helix formation,[26] GGGGSGPP, produced the opposite effect. These results are consistent with the idea that the shift from repression to induction involves a transition of at least a portion of AraC's interdomain linker from an α helix to a more flexible state, and the fraction of protein in either state can be increased by appropriate substitutions in the linker. It should be noted, however, that we have not directly shown that the linker is helical in the repressing state of AraC, only that the behavior of AraC in several types of experiments is consistent with this idea.

Our protein helical twist experiment is also consistent with the interdomain linker being an α helix during repression. If the DNA binding domain cannot freely rotate on the interdomain linker,[16] and with the knowledge that there are 3.7 residues per turn of an α helix, then deleting or inserting one residue from the presumed α-helical portion of the linker (the first six residues) will rotate the DNA binding domain by 1/3.7 of a turn, somewhat more than 90°. Such a change to both linkers in the dimeric AraC changes the relative orientations of the two DNA binding domains by ∼180°. Thus, if they had previously been antiparallel, now they will be somewhat parallel. The insertion of two residues would then bring the relative orientation between the DNA binding domains back to closer to their original relative orientation. Hence, wild type AraC and the protein containing a linker with an insertion or deletion of two residues could well repress better than protein with an insertion or deletion of a single residue. This is what was observed, thus strengthening the conclusion that at least in the region of the modifications, the linker is likely to be close to being α-helical when the protein is in its repressing state.

Our spectroscopic examination of peptides shows that the linker sequence itself, when free in solution, and thus absent any other interactions, does not show any significant helical content. However, when added to the end of a helix, it adopts a helical character emulating that of the nucleating helix itself. Thus, it is clear that at least most of the linker is capable of adopting a helical conformation when propagated from the C-terminus of such a structure. Indeed, this is precisely the environment found in the AraC protein itself, with the linker directly following the second helix of a coiled coil in the dimerization domain.[7] This provides a potential method of control of linker helicity; disrupting the helix at the base of the linker would result in the transition to more of a random coil. This notion is further supported by the data presented earlier showing that interrupting the helix by introducing a separating proline disrupts the propagation of the helix into the linker, resulting in an unstructured linker.

While this paper has primarily focused on the first six residues of the linker, it is notable that the linker ends with two proline residues. Although prolines are known to stably cap helices from N-terminal positions,[35] they have the opposite effect when located at the C-terminus and generally act to disrupt upstream helices,[36] and two consecutive proline residues are likely to even more strongly oppose helix formation. Overall, it seems possible then that several of the linker residues preceding the two proline residues at the end of the AraC linker are not truly helical. Such a structure appears to be compatible with all of the experiments we report here.

How could the structural state of the linker be controlled? As the linker is not an essential part of either domain and may extend in space away from both of them, the most apparent source of a signal to the linker is either from the N-terminal arm of AraC or directly through the residues of the dimerization domain immediately preceding the linker. Because the linker functions properly with a wide variety of sequences,[20] a direct linker−arm interaction seems unlikely. The dimerizing helix immediately precedes the linker and thus could nucleate helix formation of the linker. If arabinose controlled the helical stability or some other structural property of several of the residues of the dimerizing helix, the helical state of most of the linker could be controlled. This is an appealing mechanism because repression negative mutations lie not only in the N-terminal arm but also in the portion of the protein between the apo position of the arm and the base of the linker (Figure 6). Additionally, this is consistent with previous experiments that examined the anisotropy of the dimerization domain.[19] Thus, a potential mechanism for communication of the presence of arabinose to the linkers could be through structural changes induced by the relocation of the N-terminal arm of AraC over arabinose that then allow the residues lying between the arm and the interdomain linker to alter their conformation and thereby weaken the helix nucleating capability of the dimerizing helix.

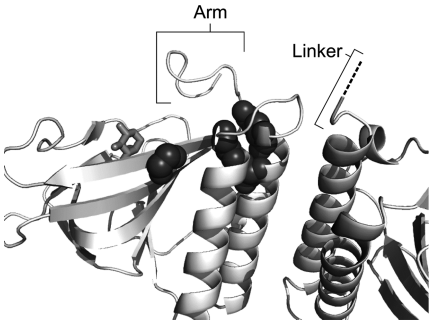



**Figure 6.** Positions of repression-disrupting mutations, G22, G141, A152, and E149, which are represented in space filling mode, in the region between the arm and linker. The apoprotein is presented with arabinose (in stick form) superimposed to illustrate its binding location.

## ■ ASSOCIATED CONTENT

### Accession Codes

AraC, CAA23508.

## ■ AUTHOR INFORMATION

### Corresponding Author

*Robert Schleif: Department of Biology, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218. E-mail: rfschleif@jhu.edu.



DOI: 10.1021/acs.biochem.9b00234
*Biochemistry* 2019, 58, 2867−2874

**ORCID**

Matthew J. Brown: 0000-0003-3640-5379
Robert Schleif: 0000-0002-4430-7085

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

Funding was provided by the Department of Biology of Johns Hopkins University.

## ■ REFERENCES

(1) Wang, X., Vallurupalli, P., Vu, A., Lee, K., Sun, S., Bai, W., Wu, C., Zhou, H., Shea, J., Kay, L. E., and Dahlquist, F. W. (2014) The linker between the dimerization and catalytic domains of the CheA histidine kinase propagates changes in structure and dynamics that are important for enzymatic activity. *Biochemistry 53*, 855−861.
(2) Ma, B., Tsai, C., Haliloğlu, T., and Nussinov, R. (2011) Dynamic allostery: Linkers are not merely flexible. *Structure 19*, 907−917.
(3) Kukic, P., Camilloni, C., Cavalli, A., and Vendruscolo, M. (2014) Determination of the individual roles of the linker residues in the interdomain motions of calmodulin using NMR chemical shifts. *J. Mol. Biol. 426*, 1826−1838.
(4) Wriggers, W., Chakravarty, S., and Jennings, P. A. (2005) Control of protein functional dynamics by peptide linkers. *Biopolymers 80*, 736−746.
(5) Sheppard, D. E., and Englesberg, E. (1967) Further evidence for positive control of the L-arabinose system by gene *araC*. *J. Mol. Biol. 25*, 443−454.
(6) Bustos, S. A., and Schleif, R. F. (1993) Functional domains of the AraC protein. *Proc. Natl. Acad. Sci. U. S. A. 90*, 5638−5642.
(7) Soisson, S. M., MacDougall-Shackleton, B., Schleif, R., and Wolberger, C. (1997) The 1.6 A crystal structure of the AraC sugar-binding and dimerization domain complexed with D-fucose. *J. Mol. Biol. 273*, 226−237.
(8) Rodgers, M. E., and Schleif, R. (2009) Solution structure of the DNA binding domain of AraC protein. *Proteins: Struct., Funct., Genet. 77*, 202−208.
(9) Dunn, T. M., Hahn, S., Ogden, S., and Schleif, R. F. (1984) An operator at −280 base pairs that is required for repression of *araBAD* operon promoter: Addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc. Natl. Acad. Sci. U. S. A. 81*, 5017−5020.
(10) Hahn, S., and Schleif, R. (1983) *In vivo* regulation of the *escherichia coli araC* promoter. *J. Bacteriol. 155*, 593−600.
(11) Carra, J. H., and Schleif, R. F. (1993) Variation of half-site organization and DNA looping by AraC protein. *EMBO J. 12*, 35−44.
(12) Martin, K., Huo, L., and Schleif, R. F. (1986) The DNA loop model for *ara* repression: AraC protein occupies the proposed loop sites *in vivo* and repression-negative mutations lie in these same sites. *Proc. Natl. Acad. Sci. U. S. A. 83*, 3654−3658.
(13) Lobell, R. B., and Schleif, R. F. (1990) DNA looping and unlooping by AraC protein. *Science 250*, 528−532.
(14) Hendrickson, W., and Schleif, R. (1985) A dimer of AraC protein contacts three adjacent major groove regions of the *araI* DNA site. *Proc. Natl. Acad. Sci. U. S. A. 82*, 3129−3133.
(15) Zhang, X., Reeder, T., and Schleif, R. (1996) Transcription activation parameters at *ara* $p_{BAD}$. *J. Mol. Biol. 258*, 14−24.
(16) Seabold, R. R., and Schleif, R. F. (1998) Apo-AraC actively seeks to loop. *J. Mol. Biol. 278*, 529−538.
(17) Harmer, T., Wu, M., and Schleif, R. (2001) The role of rigidity in DNA looping-unlooping by AraC. *Proc. Natl. Acad. Sci. U. S. A. 98*, 427−431.
(18) Weldon, J. E., Rodgers, M. E., Larkin, C., and Schleif, R. F. (2007) Structure and properties of a truely apo form of AraC dimerization domain. *Proteins: Struct., Funct., Genet. 66*, 646−654.
(19) Malaga, F., Mayberry, O., Park, D., Rodgers, M. E., Toptygin, D., and Schleif, R. F. (2016) A genetic and physical study of the interdomain linker of E. coli AraC protein-a trans-subunit communication pathway. *Proteins: Struct., Funct., Genet. 84*, 448−460.
(20) Seedorff, J., and Schleif, R. (2011) Active role of the interdomain linker of AraC. *J. Bacteriol. 193*, 5737−5746.
(21) Reed, W. L., and Schleif, R. F. (1999) Hemiplegic mutations in AraC protein. *J. Mol. Biol. 294*, 417−425.
(22) Hahn, S., Dunn, T., and Schleif, R. (1984) Upstream repression and CRP stimulation of the *escherichia coli* L-arabinose operon. *J. Mol. Biol. 180*, 61−72.
(23) Parker, A., and Gottesman, S. (2016) Small RNA regulation of TolC, the outer membrane component of bacterial multidrug transporters. *J. Bacteriol. 198*, 1101−1113.
(24) Rodgers, M. E., and Schleif, R. (2012) Heterodimers reveal that two arabinose molecules are required for the normal arabinose response of AraC. *Biochemistry 51*, 8085−8091.
(25) Schleif, R. F., and Wensink, P. C. (1981) *Practical Methods in Molecular Biology*, Springer-Verlag, New York.
(26) Pace, C. N., and Scholtz, J. M. (1998) A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J. 75*, 422−427.
(27) Berg, O., and von Hippel, P. H. (1985) Diffusion-controlled macromolecular interactions. *Annu. Rev. Biophys. Biophys. Chem. 14*, 131−160.
(28) Gutfreund, H. (1995) *Kinetics for the Life Sciences*, Cambridge University Press, Cambridge, U.K.
(29) Hendrickson, W., and Schleif, R. F. (1984) Regulation of the *escherichia coli* L-arabinose operon studied by gel electrophoresis DNA binding assay. *J. Mol. Biol. 178*, 611−628.
(30) Martin, K. J., and Schleif, R. F. (1987) Equilibrium DNA-binding of AraC protein. compensation for displaced ions. *J. Mol. Biol. 195*, 741−744.
(31) Soisson, S. M., MacDougall-Shackleton, B., Schleif, R., and Wolberger, C. (1997) Structural basis for ligand-regulated oligomerization of AraC. *Science 276*, 421−425.
(32) Timmes, A., Rodgers, M., and Schleif, R. (2004) Biochemical and physiological properties of the DNA binding domain of AraC protein. *J. Mol. Biol. 340*, 731−738.
(33) Marqusee, S., Robbins, V. H., and Baldwin, R. L. (1989) Unusually stable helix formation in short alanine-based peptides. *Proc. Natl. Acad. Sci. U. S. A. 86*, 5286−5290.
(34) Brahms, S., and Brahms, J. (1980) Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *J. Mol. Biol. 138*, 149−178.
(35) Presta, L. G., and Rose, G. D. (1988) Helix signals in proteins. *Science 240*, 1632.
(36) Kim, M. K., and Kang, Y. K. (1999) Positional preference of proline in α-helices. *Protein Sci. 8*, 1492−1499.



DOI: 10.1021/acs.biochem.9b00234
*Biochemistry* 2019, 58, 2867−2874