Published in final edited form as:
Proteins. 2009 January ; 74(1): 81–91. doi:10.1002/prot.22137.

Functional Modes of The Regulatory Arm of AraC

Michael E. Rodgers, **Nakisha D. Holder**, **Stephanie Dirla**, and **Robert Schleif**
Biology Department, Johns Hopkins University, Baltimore, MD

Abstract

One of the two crystal structures of the arm-dimerization domain determined in the absence of arbinose fails to show the arm, whereas the other structure does show it. The two structures lead to different pictures for the regulatory behavior of the arms. Trypsin digestion, fluorescence anisotropy, and NMR experiments presented here were designed to resolve the issue and show that in arm-dimerization domain, the arms are structured, although differently, in the presence and absence of arabinose. The arms have also been shown to interact with the DNA binding domains of AraC by their requirement for the immobilization of the DNA binding domains that is necessary for DNA looping and repression. The binding of arabinose has been shown to release the DNA binding domains and looping ceases. The picture resulting from the new experiments and the crystal structures of the arm-dimerization domain is that in the absence of arabinose, the arm adopts one structure on the dimerization domain and that the DNA binding domain then binds to this complex. Upon binding arabinose, the arm restructures and as a result, no longer serves as a gasket between the DNA binding domain and dimerization domain. The DNA binding domain is then released, subject only to the constraints imposed by the flexible linker connecting it to dimerization domain, and the protein relocates on the DNA and activates transcription.

Keywords

Peptidyl arm; gasket; domain immobilization; NMR-HSQC; AraC; gene regulation; transcription

INTRODUCTION

Residues 7–17, the N-terminal arms of the arabinose operon regulatory protein in *Escherichia coli*, AraC, play a critical role in controlling whether the protein represses or induces expression of the arabinose catabolic genes in *Escherichia coli*[1–4]. These residues are well structured and lie over bound arabinose in crystals of the arm-dimerization domain (arm is residues 1–18, dimerization domain is residues 20–166) of AraC formed in the presence of arabinose[5]. X-ray diffraction data from arm-dimerization domain crystals formed in the absence of arabinose however, lacks electron density for the arm[5], implying that the arm is unstructured and does not bind to the dimerization domain in this condition. This picture of the arms binding to the dimerization domain only when arabinose is present has been implicitly incorporated into the light switch mechanism that summarizes data contained in the above-mentioned papers, plus a substantial amount of data subsequently collected[6–10].

In addition to its interaction with the dimerization domain in the presence of arabinose, a number of genetics studies strongly suggest that in the absence of arabinose, the N-terminal arms of AraC interact with the protein's DNA binding domains, DBDs[2,3,9]. The simplest

Corresponding author: Robert Schleif, Biology Department, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD21218, Phone: 410 516-5206, Fax: 410 516-5213, Email: schleif@jhu.edu.

mechanism consistent with the data is that in the absence of arabinose, the arms, although attached to the dimerization domain, are not bound or structured by it. This leaves them available to bind to the DBDs. As a result of the arm-DBD interaction, AraC is held in a structure that favors DNA looping as represented in Fig. 1. In these repression loops, one DBD binds to the $araI_1$ half- site and the other binds to the $araO_2$ half-site located 210 base pairs away. In the presence of arabinose, a stronger interaction between the arm and dimerization domain favors a shift of the arms from the DBDs to the dimerization domains, thereby freeing the DBDs from the arm constraint. Considerable evidence has been generated supporting this light switch model. The mechanism as described however, leaves several questions unanswered. First, it is not at all clear how the binding of a part of the N-terminal arm to the DBD immobilizes the domain so that DNA looping is preferred. Second, in crystals of a nonaggregating mutant of the arm-dimerization domain, the arm was seen to be structured, but in a different conformation from that seen in the presence of arabinose11. Because residue 12 of the arm in this new structure participates in crystal lattice contacts, the significance of the structuring of the arm in the mutant, which is seen in the absence of the DBD, cannot be ascertained.

In the work described here, we have investigated the structural properties of the N-terminal arm of AraC and have attempted to answer the two questions raised above concerning the arm. We provide evidence that the arm itself is unstructured in solution, but that it is structured when it is in its natural position linked to the beginning of the dimerization domain. Furthermore, the arm is not structured in the presence of just the DBD. Thus, in the absence of arabinose, arm is structured by its interactions with the dimerization domain, and the DBD is then immobilized by its interaction with the arm-dimerization domain. It is also likely that the structure of the arm as seen in the mutant arm-dimerization domain crystals represents the structure to which the DBDs of AraC bind in the absence of arabinose. The binding of arabinose leads to restructuring of the N-terminal arm and release of the DBDs. This allows AraC protein to bind to the adjacent $araI_1$-$araI_2$ half-sites and activate transcription from the *ara* p_{BAD} promoter.

METHODS

Expression vectors for arm-dimerization domain, DBD, and N-terminal arm-linker(s)-DBD

The arm-dimerization domain expression vector, AraCTF, has been previously described12. Essentially, the AraCTF plasmid was derived from pET21b by cloning nucleotides 1–546 of AraC between the *Nde*I and *Ade*I sites. The resulting vector expresses residues 1–182 of AraC, which corresponds to the arm, dimerization domain, and inter domain linker regions of the protein followed by a C-terminal Leu-Glu linker and a His-6 tag. The mutations Q5C, L10D, N17R, Y31V, and C66S were introduced with Quick-change mutagenesis (Stratagene) and verified with DNA sequencing. The DNA encoding the AraC DBD region, amino acid residues 175–292, was inserted between the *Nco*1 and *Sac*1 sites of the multiple cloning region in the pET24d vector. The DNA encoding the native arm (residues 1–17), flexible linkers (amino acid sequences GNATA, GNATAPDVSQ, and GNATAPDVSQGTAIS) and native linker (amino acid sequence NESLHPP) were added in several steps of Quick-change mutagenesis and verified by DNA sequencing.

Purification and isotopic labeling of dimerization domain

Arm-dimerization domain protein was expressed from the AraCTF plasmid in BL21 (DE3) and purified essentially as described previously11. Cells were grown at 37°C in YT medium13 containing 200 μg/mL ampicillin to an apparent $OD_{550\ nm}$ of 0.8 to 1.0, induced with the addition of IPTG to 0.8 mM for a minimum of three hours, then harvested by centrifugation. For NMR studies the mutant Y31V11 mutant arm–dimerization domain was

used. ^{15}N-labeled protein was over-expressed using a modification of the method of Marley14 where growth prior to the labeling step was carried out in minimal medium rather than in rich medium.

Cells were lysed by a French press in 15 mM Tris-Cl, pH 8.0, 0.1 M NaCl, 5% glycerol, 10 mM $MgCl_2$, 50 mM L-arabinose with 10 mg/mL each DNase I and RNase A added immediately before use, or by grinding with alumina13, or by digestion with 0.4 mg lysozyme/gm cells in the presence of 0.06% deoxycholate for 1.5 hr. After centrifugation, the supernatant from the lysed cells was mixed with 1 ml Ni-NTA agarose beads from Qiagen (Valencia, CA) per estimated 10 mg target protein and rocked gently at 4°C for a minimum of 2 h. Typically, 5 ml of beads was incubated with extract from the cells obtained from 1 l of cell growth medium. Unbound material was removed and the beads were washed with 15 mM Tris-Cl, pH 8.0, 0.1 M NaCl, 50 mM L-arabinose, and 10 mM imidazole until the A_{280} nm of the wash was <0.05. Bound protein was eluted with two column volumes of buffer containing 15 mM Tris-Cl, pH 8.0, 10 mM NaCl, 50 mM L-arabinose, 1 M imidazole. For NMR studies the protein was digested with 0.01 mg trypsin per 1 mg estimated His-6 protein overnight at 4°C to release the His-6 tag. This was followed by anion exchange chromatography on a 1 ml Pharmacia Mono-Q HR 5/5, 15 mM Tris-Cl, pH 8.0, 50 mM L-arabinose using an elution gradient of 10 mM to 1 M NaCl over 40 ml.

Purification and isotopic labeling of DBD and N-terminal arm+linker(s)+DBD

An isolated colony of cells, BL21 (DE3), freshly transformed with the appropriate pET24d expression construct was grown for 8–16 hours in 5 ml YT medium. 40 μg/ml kanamycin was used in all growth media. Unlabeled proteins were over expressed for 3 hours in cultures grown in YT and induced with 0.4 mM IPTG once the A_{550} reached 0.8. 15N-labeled proteins were over-expressed as described above.

The purification of both DBD and various arm-linker-DBD constructs are essentially identical. Frozen cells (typically about 5g) were thawed on ice, resuspended in 3 volumes of lysis buffer, 20 mM sodium phosphate, pH 6.0, 500 mM NaCl, 1 mM EDTA, 1 mM DTE, 5% glycerol, and lysed by two passages through a French press. 1/100 volume of 0.1 M PMSF, freshly dissolved in ethanol, was added immediately before and again after lysis. The lysate was diluted with an equal volume of lysis buffer and centrifuged for ten minutes at 12,000 rpm in a Sorvall SS-34 rotor. The supernatant was discarded and the pellet was resuspended in 3 volumes of 7 M GuHCl and 10 mM DTE, diluted in half with 50 mM NaCl, 20 mM Na phosphate, pH 6 and dialyzed twice into 1 l of the same buffer. The dialysate was centrifuged for ten minutes at 12,000 rpm in a Sorvall SS-34 rotor and the supernatant loaded onto a 5 ml Heparin Hi-Trap column (Pharmacia) pre-equilibrated with 20 mM Na Phosphate, pH 6. Protein was eluted from the column with a 90 ml, 0–1 M NaCl gradient. The desired protein is the only significant elution peak and is >95% pure based on SDS-PAGE. For highest purity, the pooled peak fractions from the Heparin HiTrap column were further chromatographed on a Mono-S column using the same buffers and protocol. Heavily overloaded samples run on 15% SDS polyacrylamide gels showed no trace of contaminating proteins. Typically, about 50–100 mg of purified DBD are obtained from 5 g of cells. Purified protein concentrations were calculated from A_{280} using an extinction coefficient of 8,480 $M^{-1}cm^{-1}$ for DBD and 9,970 $M^{-1}cm^{-1}$ for N-term arm-linker-DBD and N-term arm-linker-native linker-DBD as calculated from the amino acid content by the method of Pace15. Since both proteins were refolded out of GuHCl, the fraction of active protein was assessed using a fluorimetric DNA binding assay. These proteins are typically fully active within experimental error.

Arabinose isomerase assay

Cells were grown in M10 minimal medium containing 0.8% glycerol, 1% casamino acids, 10 μg/ml thiamine, 20 μM $CaCl_2$, 10 μM $MgCl_2$, 100 μg/ml ampicillin with or without 0.2% L-arabinose. Cells were grown exponentially for at least eight doublings and harvested at an apparent OD600 of 0.3 to 0.9. Arabinose isomerase was assayed as described previously13. Uncertainties in cellular enzyme levels measured by this method are dominated by fluctuations in the colorometric measurement of the ribulose product of the arabinose isomerase reaction. Measurements were made in duplicate or triplicate and are about 20% of the measured level, becoming larger when the ribulose level is low.

Trypsin digestion experiments

0.06 mg/ml purified arm-dimerization domain was digested with 0.01 mg/ml trypsin in 100 μl 0.01 M Tris-HCl pH 7.5, 0.1 M NaCl, 2 mM dithiothreitol, 0.1 mM EDTA. Arabinose, if present, was at 10 mM. At the indicated times, 8 μl aliquots were removed and added to 2 μl of 0.3 M Tris-base, 10% SDS, 3.5 M 2-mercaptoethanol, 50% glycerol, 0.5% bromophenol blue, 6 mM phenylmethylsulphonyl fluoride, and then subjected to electrophoresis on 14% SDS gels, and silver stained.

MALDI-TOF mass spectrometry

Samples of dimerization domain, 30 μl, digested with trypsin as described above for 10, 30, and 90 min were exchanged into into 0.1% trifluoroacetic acid by passing through a 250 μl G-10 Sephadex gel filtration column. 10–100 picomoles of protein were spotted on an Applied Biosystems Voyager Sample Plate with sinapinic acid in 50% acetonitrile and 0.05% TFA used as the matrix. Analysis was done on a Applied Biosystems Voyager DE-STR MALDI-TOF instrument at the Mass Spectrometry Facility at the Johns Hopkins School of Medicine.

Fluorescent labeling

AraC arm-dimerization domain was labeled with fluorescein-5-isothiocyanate under conditions favoring modification of the N-terminal amine. Solution containing the fluorescent probe, 100 μl at 10 mg/ml in N, N-dimethyl formamide was added to 1 ml of 5 mM protein in 100 mM sodium bicarbonate, pH 7.4, 150 mM NaCl, and incubated at room temperature with gentle rocking for 1 hour. The labeled protein mixture was separated from unreacted label by passage over a 10 ml Sephadex G-25 column. These experiments used the Q5C N17R C66S mutant protein. We tested that the majority of the fluorescein labeling of the protein was at the N-terminus by digestion with trypsin. SDS gel electrophoresis verified that the arm was cleaved after residue R16 without cleaving elsewhere in the dimerization domain. The digestion reduced anisotropy from 0.15 to 0.07, consistent with release of labeled arm from the protein.

Fluorescence anisotropy measurements

Fluorescence anisotropy measurements were made on 120 nM protein in 2.5 ml of 20 mM Tris-HCl pH 7.4, 150 mM NaCl in 1 cm path length cuvettes. Parallel and perpendicular emission components were measured at 520 nm with vertical excitation at 494 nm from a 75 watt xenon light source. Monochrometer slits were set to 2.5 mm corresponding to a spectral band width of 10 nm. Each point in a titration was measured three times, collecting data for 10 seconds each time. Three minutes were allowed for equilibration after adding new sample and mixing.

araI$_1$ half-site DNA

A 27 bp DNA containing the 17 base sequence of the *araI$_1$* half-site and 5 base extensions on both ends was obtained from Integrated DNA Technologies, Inc. The recognition strand sequence is CGCCATAGCATTTTTATCCATAAGATC with the I_1 site underlined. The recognition strand and its complement were purified by HPLC, combined stoichiometrically and annealed. The dsDNA was further purified by HPLC to separate any excess ssDNA. Concentrations were determined using A_{260} measurements. Extinction coefficient for the recognition strand, complement and dsDNA were 216225 $M^{-1}cm^{-1}$, 226544 $M^{-1}cm^{-1}$ and 418195 $M^{-1}cm^{-1}$, respectively. Extinction coefficients were determined using A_{260} measurements before and after complete phosphodiesterase I digestion. The extinction coefficient of a digested sample was calculated as a linear combination of the respective deoxynucleotide extinction coefficients times their occurrence in the sequence.

NMR

^{15}N-labeled proteins were prepared as described above. NMR samples were prepared by dialyzing the proteins into 444.4 mM NaCl, 55.5 mM Na phosphate, pH 6. After dialysis, D_2O was added to 10% (v/v) giving a final composition of 400 mM NaCl, 50 mM Na phosphate. If necessary, samples were concentrated using a Millipore Centricon YM-10 concentrator. Typical final protein concentrations were between 200 and 600 μM. In experiments with I_1-DNA, the salt concentration was reduced to 200 mM final to insure full binding. All NMR data were collected at 25 °C.

The ^1H-^{15}N HSQC experiments were carried out on various spectrometers operated by the Biomolecular NMR Center at the Johns Hopkins University as indicated in the figure legends. Most were done on a Varian INOVA spectrometer operating at a ^1H Larmor frequency of 500 MHz. The ^1H-^{15}N HSQC spectrum of the DNA complex was obtained on a Varian INOVA spectrometer operating at a ^1H Larmor frequency of 800 MHz and equipped with a cryogenic probe with z-axis pulsed field gradients. The dimerization domain experiments were carried out on a Bruker Avance spectrometer operating at a ^1H Larmor frequency of 600 MHz and equipped with a TCI cryoprobe with z-axis pulsed field gradients. All data were processed using NMR Pipe16 and analyzed using NMRView17.

The cross-peaks in the HSCQ spectrum of DBD were assigned by using four triple resonance experiments (HNCA, HNCACB, HN(CO)CA and CBCA(CO)NH, data not shown) collected on a ^{13}C,^{15}N doubly labeled sample. Similarly, the arm, linker and DBD residue peaks were assigned on a ^{13}C,^{15}N doubly labeled arm+5 residue linker+DBD sample using the same types of experiments (data not shown). The assigned ^1H-^{15}N HSQC cross-peaks for the arm residues were found to exactly superimpose in all arm+linker+DBD samples having 5, 10, 15 and 22 residue linker lengths. Also, the DBD cross-peaks from residues 182 through 292 (AraC numbering) exactly superimpose in all constructs. The superposition of cross-peaks from a free arm peptide with the corresponding cross-peaks in arm+linker+DBD was confirmed by tryptic proteolysis of the later and obtaining ^{15}N-HSQC spectra on the resulting peptides. Complete purification of the peptides was not achieved but one peptide sample showed all of the expected arm cross-peaks.

RESULTS

Trypsin susceptibility experiments

As mentioned in the introduction, X-ray structure data is equivocal as to the status of the N-terminal regulatory arm of AraC in the absence of arabinose and the absence of the DBD. We therefore examined the structural status of the arm under these two conditions by its sensitivity to trypsin cleavage. When the arm is stably structured, it should be less sensitive

to trypsin digestion than when it is either unstructured or its stability in the structured state is reduced.

Lacking an arginine or lysine residue, the wild type N-terminal arm of AraC is not cleaved by trypsin. Although the activity of AraC is sensitive to the identities of most residues in the regulatory arm beyond position 74, we found that the N16R mutation, which introduces a trypsin cleavage site, does not significantly alter the inducing or repressing activities of AraC, Table 1. To facilitate other anticipated studies on the protein, the mutations Q5C and C66S were also added

Trypsin digestion followed by electrophoresis on SDS gels showed that within 15 minutes of digestion, Fig. 2, the C-terminal tail on the protein is cleaved. whereas after 75 minutes of digestion, the amount of armless wild type dimerization domain was only slightly increased above the amount present at the beginning of the experiment. Mass spectrometry of the same samples confirmed that the initial cleavage occured at residue R178 and that the slower cleavage occurred after the engineered residue N17R. Dimerization domain beyond residue 167 is not structured5, and should be cleaved rapidly by trypsin. The fact that the trypsin cleavage site in the arm of arm-dimerization domain was so much less sensitive to trypsin suggests that it is structured and not readily accessible to the trypsin. In this and other experiments the difference in digestion rate plus and minus arabinose was barely discernable. These results are consistent with the arm, when present on the dimerization domain, possessing structure both in the presence and absence of arabinose rather than being structured only when arabinose is present.

The properties of constitutive mutations in AraC can provide additional evidence that the arm is normally structured on the dimerization domain in the absence of arabinose and that the protease sensitivity experiment is responsive to the structure of the arm. A constitutive AraC mutant activates transcription without the need for arabinose. Deletions of the regulatory arm and most point mutations in the arm are constitutive2‚4‚6, apparently resulting from an inability of the arm in these mutants to bind the DBD and leaving the protein by default to activate transcription. Some constitutive mutations in the arm-dimerization domain should leave the arm unstructured or reduce the stability of its structure. This is, in fact the case, as shown in the bottom half of Fig. 2. The trypsin sensitivity of the arm in a constitutive mutant, L10D, was increased three to five-fold in the presence or absence of arabinose compared to the sensitivity of wild type protein.

Fluorescence anisotropy measurements

The fluorescence anisotropy of a label on the arm on the arm-dimerization domain could provide evidence supporting the possibility raised by the trypsin digestion experiments. If the arm on the arm-dimerization domain were to become more structured upon the binding of arabinose, the anisotropy would increase. We therefore labeled the N-terminus of arm-dimerization domain with fluorescein isothiocyanate. Fig. 3 shows the anisotropy as a function of the concentration of added glucose, which should produce little or no response, and arabinose. The anisotropy decreased upon arabinose addition, indicating that the end of the arm became more free to tumble upon arabinose addition rather than less free. Although the relevant portion of the arm is residues 7–17, this result, like the trypsin result, suggests that the arm is structured in the absence of arabinose and that it becomes somewhat less structured in the presence of arabinose. The finding of a decrease in the anisotropy upon arabinose addition could also be a result of the dimerization domain becoming more symmetrical upon arabinose addition. Such an effect seems unlikely to contribute significantly to the anisotropy change because the structure of the dimer, exclusive of the N-terminal arms, is not significantly different in the two states5‚11.

NMR experiments with arm-dimerization domain

The trypsin digestion and fluorescence anisotropy experiments give little direct information on the extent of the arm that may be structured or on the fraction of time that the N-terminal arm may be structured. NMR spectroscopy however, is capable of providing this information. The ^{1}H-^{15}N HSQC spectra of smaller proteins usually display well-separated spectral cross-peaks for most of the amide protons. Some of these cross-peaks shift if the environment of the amides is altered, for example by the binding of a peptide nearby. In addition, the HSQC cross-peaks from unstructured regions, as might exist for an unbound or dynamically flexible peptide like the N-terminal arm of AraC, are often sharper and of considerably greater amplitude than cross-peaks from amides in structured regions.

The molecular weight of a dimer of the dimerization domain, 37,000, is too high for the ^{1}H-^{15}N HSQC spectrum to display well-resolved amide hydrogen cross-peaks from structured regions of the protein. Residues in unstructured regions however, would still be expected to produce intense, sharp cross-peaks. Fig. 4 shows an ^{1}H-^{15}N HSQC spectrum of residues 2–166 of AraC taken in the presence and the absence of arabinose and contoured to show the most intense peaks. None of the few intense and sharp peaks derive from backbone amides of unstructured arm. The positions of such peaks as determined from arm+linker +DBD constructs and confirmed in a separate experiment with free arm, are shown by the stars. Signals from Pro8 and Pro11 of the arm would not be seen because they lack an amide hydrogen. The handful of peaks that are seen likely derive from loop regions in the dimerization domain. The results also show that few of the cross-peaks seen in the ^{1}H-^{15}N HSQC spectrum shift in position on the addition of arabinose. This is not the behavior that would be expected if the cross-peaks derived from residues of the arm that became structured in the presence of arabinose. In addition, the number of cross-peaks deriving from unstructured regions of the protein did not significantly change upon arabinose addition.

The arm does not bind to free DBD

The experiments described above showed that the N-terminal regulatory arm on AraC is stably structured on the dimerization domain in the absence of the DBD, both in the presence and absence of arabinose. Is the arm also structured by its interaction with the DBD? This question can also be answered with ^{1}H-^{15}N HSQC NMR experiments. The first question we addressed is whether or not free N-terminal arm, that is residues 1–17, can bind to the DBD when the arm and DBD are both in solution.

Initial experiments showed no detectable arm-DBD interaction. This is not surprising because the interaction between the two should, in fact, be weak. In the native AraC protein, the arms are held within 50 Å of a DBD. Thus, the effective concentration of an arm near a DBD is one molecule in a sphere of 50 Å radius, equivalent to a solution concentration of about 0.02 M. This is well above the peptide and protein concentrations feasible in NMR measurements with these proteins. To overcome the concentration limitation, we engineered constructs in which the first 17 residues of the N-terminal arm of AraC were connected to the N-terminus of the DBD through unstructured peptide linkers of lengths 5, 10, 15, and 22 residues. A 17-residue arm was chosen because it had previously been shown to exhibit arabinose-dependent binding to dimerization domain, mimicking wild-type function8. As each engineered linker plus arm was connected to the DBD at the same point as the interdomain linker in full length AraC protein, the addition of an arm plus linker should have no deleterious effects on the structure or function. The NMR experiments described below confirm this as do DNA binding measurements. DBD and the arm-linker-DBDs have similar binding affinities for I_1 DNA, (data not shown).

Fig. 5 shows, at two contour levels, ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled DBD with an arm attached via a 22-residue linker. Although the arm residues and the entire DBD have been assigned, the peaks from only the arm are identified in the figure. The cross-peaks deriving from residues in the arm are all intense and sharp, indicating that the added linker plus arm does not possess a stable, static structure. Most of the additional intense and sharp cross-peaks in the spectrum derive from the last 10 residues at the C-terminus, indicating that the C-terminal tail is also unstructured. Except for a few cross-peaks at the connection point, the cross-peaks from the DBD in the arm-linker-DBD construct superimpose with the cross-peaks in DBD alone, also showing that there is no significant binding interaction between the arm and the DBD. We conclude that the N-terminal arm is unable to bind to the DBD.

DBDs of proteins frequently undergo appreciable change in structure upon binding to DNA, most often by becoming more structured. Therefore, it is possible that the DBD of AraC is capable of interacting with the N-terminal arm only when the domain is bound to DNA. To test this possibility, we performed the NMR measurements in the presence of an oligonucleotide containing the AraC I_1 DNA binding half-site. The affinity of DBD for the I_1 half-site, K_d = 120 nM in 200 mM KCl18, ensures that more than 98% of the DBD will be bound to the DNA at the protein and DNA concentrations used in the experiment. The peaks deriving from the arm residues in an ^{1}H-^{15}N HSQC spectrum of the arm+22-residue linker-DBD were intense and sharp and not significantly changed in position (data not shown). Thus, the ^{1}H-^{15}N HSQC spectra of the DBD, either free or while bound to DNA, showed that the N-terminal arm in the absence of the dimerization domain does not bind with any significant affinity to the DBD.

Apo and plus arabinose arm crystal structures significantly differ

The first crystal structures obtained of the arm-dimerization domain of AraC were of the protein with bound arabinose and of the apo protein lacking arabinose5. The structure resulting from crystallization in the absence of arabinose did not reveal the arm. A more recently obtained structure from protein crystallized in the absence of arabinose did show the arm11. The two arm-dimerization domain structures for which the arm is visible, that with arabinose bound, and that with no arabinose, were virtually identical except for the structures of the arms, where they differed significantly, Fig. 6a. The different appearance might be trivial, for example the shifting of just a couple polypeptide backbone phi and psi values. Fig. 6b shows this is not the case and that these angles substantially differ between the two structures for most of the residues between position 7 and 18. There is also a change in residue 22, which acts to slightly tilt the end of the first β-strand.

DISCUSSION

Data from a wide variety of experimental approaches has lead to a picture for the mechanism by which AraC protein shifts from a repressing state to a state in which the protein activates transcription. The original structures of the N-terminal arm-dimerization domain of AraC crystalized in the presence and absence of arabinose5 importantly contributed to the model1⁻4. In crystals of wild type arm-dimerization domain prepared in the absence of arabinose, electron density for the arm was not visible, indicating that it was not structured. In crystals of wild type protein prepared in the presence of arabinose however, the arm was observed and lay over the bound arabinose5. The simplest model consistent with this structure data and additional genetic, physiological, and biophysical data is that the arm itself binds to the DBD in the absence of arabinose6⁻10. This binding sufficiently immobilizes the DBDs that DNA looping is preferred. Upon the addition of arabinose, the arms bind to the dimerization domains over the arabinose, thereby releasing the DBDs and allowing them to bind to the adjacent half-sites beside the RNA polymerase binding site. When bound to these sites, AraC stimulates the initiation of transcription of the

ara operon genes. Deeper examination of the properties of AraC has provided a number of clues how the light switch mechanism as outlined above for AraC might be extended. As explained below, the clues are not definitive. The objective of the work described in this paper was to provide stronger evidence for the extension that is pointed to by the clues.

First, in the crystals of wild type arm-dimerization domain containing arabinose, the space above the arabinose binding pocket is occupied by the N-terminal arm. In crystals of wild type arm-dimerization domain lacking arabinose, this space is occupied by another dimerization domain that binds to the first dimer in a face-to-face manner5. Thus, it is possible that in solution, in the absence of arabinose, the arm in wild type protein normally binds near the pocket but that it is displaced upon crystallization by the face-to-face interaction with another dimerization domain.

Second, while the face-to-face interaction of arm-dimerization domain occurs at moderate concentrations in solution in the absence of arabinose, it occurs at much lower protein concentrations when the N-terminal arm is mutated or deleted19. This suggests that in the absence of arabinose, the arm in wild type protein binds to a site on the dimerization domain from which it must be displaced in order that the face-to-face interaction occur. This result cannot be definitive however, due to the possibility of competing effects such as the arm acting as a solubility life raft, *i.e*, something which merely acts to increase solubility.

Third, the availability of a *trans*-complementing peptide sequence fused to the N-terminus of the N-terminal arm of arm-dimerization domain was measured by its ability to complement inactive ω-deleted β-galactosidase. The complementing activity was not reduced upon arabinose addition, and in fact, was slightly increased7. These results suggest that the native AraC arm binds to the dimerization domain in the absence of arabinose and that it becomes somewhat more free in the presence of arabinose. It is also possible, however, that the added peptide sequence itself binds to the dimerization domain, and that forming the plus arabinose structure of the arm freed the added peptide sequence from its binding site.

Fourth, to facilitate physical studies, the interaction primarily responsible for the face-to-face interaction of arm-dimerization domain in the absence of arabinose was eliminated by mutation. This allowed the mutant protein to crystallize without the face-to-face interaction seen with the wild type protein11. The fact that X-ray diffraction from these crystals clearly delineated the N-terminal arms from residue 7 onwards appears to demonstrate that the arm in wild dimerization domain is structured on the dimerization domain in the absence of arabinose. Unfortunately, it is also possible that a crystal lattice contact involving residue 12 of the arm might have been mainly responsible for structuring of the arm.

In the experiments described in this paper, we provide more definitive evidence than the hints mentioned above that the arm is in fact structured on the dimerization domain both in the presence and absence of arabinose. In some of the experiments described here, we introduced the N16R mutation in the regulatory arm to add a trypsin cleavage site to the arm without altering the regulatory properties of AraC. The cleavage rate at the new trypsin site in arm-dimerization domain constructs was very low compared to cleavage at an unstructured site at the C-terminal end of the protein. Furthermore, the cleavage rate was was only moderately altered by the presence of arabinose, suggesting that the arm's structural stability is nearly the same in the presence and absence of arabinose. Additionally, since the structure data shows that the arm is structured in the presence of arabinose, the trypsin experiments indicate that the arm is also structured in the absence of arabinose. To verify that the cleavage rate of the N16R mutant arm is sensitive to structure, we introduced a change in the arm sequence whose physiological behavior indicates that the arm no longer

interacts with the DBD, presumably because the mutant arm's interaction with the dimerization domain has been weakened or destabilized. The trypsin cleavage rate of the mutant arm was substantially increased.

If the arm on the arm-dimerization domain were to become structured by the binding of arabinose, then the addition of arabinose ought to reduce tumbling rate of the N-terminal residue of the arm. The tumbling was monitored by linking fluorescein to the N-terminus and measuring fluoresence anisotropy. Instead of increasing upon arabinose addition, the anisotropy slightly decreased, a result consistent with the trypsin results. This result however, does not definitively show that the native arm is structured in the absence of arabinose because the fluorescein could have bound to a hydrophobic patch on the dimerization domain, and have been unable to bind to the patch in the presence of arabinose.

To provide stronger evidence that the arm is stably structured on the dimerization domain with or without arabinose, we examined the structural status of the arm by NMR. These experiments showed that the arm itself is unstructured in solution, but that it is structured when it is connected in its normal position to the dimerization domain. Further, it is structured both in the presence and absence of arabinose. These results suggest that the structuring of the arm that is seen in the crystals obtained in the absence of arabinose is unlikely to be the result of crystal lattice contacts.

In previously reported plasmon resonance experiments investigating the binding of N-terminal arm peptide to arm-deleted dimerization domain, we found detectable interactions between arm and dimerization domain only in the presence of arabinose8. One might expect on the basis of the results presented here that an interaction would have been detected both in the presence and absence of arabinose. Since, however, the plus arabinose interaction must be substantially stronger than the minus arabinose interaction in order to shift most of the protein in the cell from the repressing state to the inducing state, it is understandable that only the plus arabinose interaction was detected.

The allosteric mechanism displayed by AraC protein in changing its DNA binding properties in response to the binding of the small molecule ligand arabinose does not appear to have been observed in other proteins. This paucity of mechanistically similar examples may derive more from the scarcity of systems whose details of action have been determined than from nature's infrequent use of the mechanism. Another reason may be that inappropriate screening methods have been used. The version of the light switch mechanism extant before this work suggested that other proteins possessing the same regulatory mechanism could simply be identified by the effects of short N-terminal deletions upon their regulatory properties—in the case of AraC, by becoming constitutive. In light of the new results presented here, such a simple identification may not generally be possible. Since the N-terminal arm of AraC appears to be structured on the dimerization domain both in the presence and absence of arabinose, slightly different energetics could have resulted in the requirement that the full arm always be present in order for the protein to fold or to be stable. In fact, markedly different stabilities are even observed for AraC with different N-terminal deletions1. Thus, the identification of other proteins utilizing the same basic mechanism as AraC may require physical characterization rather than characterization of the regulatory properties of truncated proteins. Overall, it seems likely that some of the other members of the large AraC family of homologs20 will turn out to use some of the same regulatory principles as are utilized by AraC.

CONCLUSION

The results of the NMR experiments indicate that the N-terminal regulatory arm folds into a stable structure in conjunction with the dimerization domain. As shown earlier6, the arm is required for immobilization of the DBD. Thus, the DBD must then bind to the structured arm-dimerization domain. Hence, the arm appears to serve as a necessary gasket between the dimerization and DBDs. Upon the addition of arabinose, the arm restructures and no longer serves as a gasket. The DBD is thus freed from the arm constraint and is free to reposition and reorient, limited only by the interdomain linker between the dimerization and DBDs.

Acknowledgments

We thank Dr. Ananya Majumdar and the Johns Hopkins Biomolecular NMR Center for assistance with all NMR experiments and for critical reading of the manuscript. We also thank Katie Frato, Victoria Hargreaves, and Jennifer Seedorff for comments on the manuscript and NIH grant GM18277 for support.

References

1. Saviola B, Seabold R, Schleif R. Arm-domain interactions in AraC. J Mol Biol. 1998; 278:539–548. [PubMed: 9600837]
2. Seabold R, Schleif R. Apo-AraC actively seeks to loop. J Mol Biol. 1998; 278:529–538. [PubMed: 9600836]
3. Reed W, Schleif R. Hemiplegic mutations in AraC protein. J Mol Biol. 1999; 294:417–425. [PubMed: 10610768]
4. Ross J, Gryczynski U, Schleif R. Mutational analysis of residue roles in AraC function. J Mol Biol. 2003; 328:85–93. [PubMed: 12683999]
5. Soisson S, MacDougal-Shackleton B, Schleif R, Wolberger C. Structural basis for ligand-regulated oligomerization of AraC. Science. 1997; 276:421–425. [PubMed: 9103202]
6. Schleif R. The AraC protein: a love-hate relationship. BioEssays. 2003; 25:274–282. [PubMed: 12596232]
7. Gryczynski U, Schleif R. A portable allosteric mechanism. proteins. 2004; 57:9–11. [PubMed: 15326589]
8. Ghosh M, Schleif R. Biophysical evidence of arm-domain interactions in AraC. Analyt Biochem. 2001; 295:107–112. [PubMed: 11476551]
9. Wu M, Schleif R. Mapping arm-DNA-binding domain interactions in AraC. J Mol Biol. 2001; 307:1001–1009. [PubMed: 11286551]
10. Harmer T, Wu M, Schleif R. The Role of rigidity in DNA looping-unlooping by AraC. Proc Natl Acad Sci USA. 2001; 98:427–431. [PubMed: 11209047]
11. Weldon J, Rodgers M, Larkin C, Schleif R. Structure and properties of a truely apo form of AraC dimerization domain. Proteins. 2007; 66:646–654. [PubMed: 17173282]
12. LaRonde-LeBlanc N, Wolberger C. Characterization of the oligomeric states of wild type and mutant. AraC Biochemistry. 2000; 39:11593–11601.
13. Schleif, R.; Wensink, P. Practical methods in molecular biology. Springer-Verlag; New York: 1981.
14. Marley J, Lu M, Bracken C. A method for efficient isotopic labeling of recombinant proteins. J Biomol NMR. 2001; 20:71–75. [PubMed: 11430757]
15. Pace CN, Vajdos F, Lee L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. Protein Sci. 1995; 4:2411–2423. [PubMed: 8563639]
16. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995; 6:277–293. [PubMed: 8520220]
17. Johnson BA, Blevins C. NMRView: A computer program for the visualization and analysis of NMR data. J Biomol NMR. 1994; 4:603–614.

18. Timmes A, Rodgers M, Schleif R. Biochemical and physiological properties of the DNA binding domain of AraC protein. J Mol Bio. 2004; 340:731–738. [PubMed: 15223316]

19. Weldon J, Schleif R. Specific interactions by the N-terminal arm inhibit self-association of the AraC dimerization domain. Protein Science. 2006; 15:2828–2835. [PubMed: 17132863]

20. Gallegos M, Schleif R, Bairoch A, Hofmann J, Ramos J. AraC/XylS Family of Transcriptional Regulators. Microbiology and Molecular Biology Reviews. 1997; 61:393–410. [PubMed: 9409145]

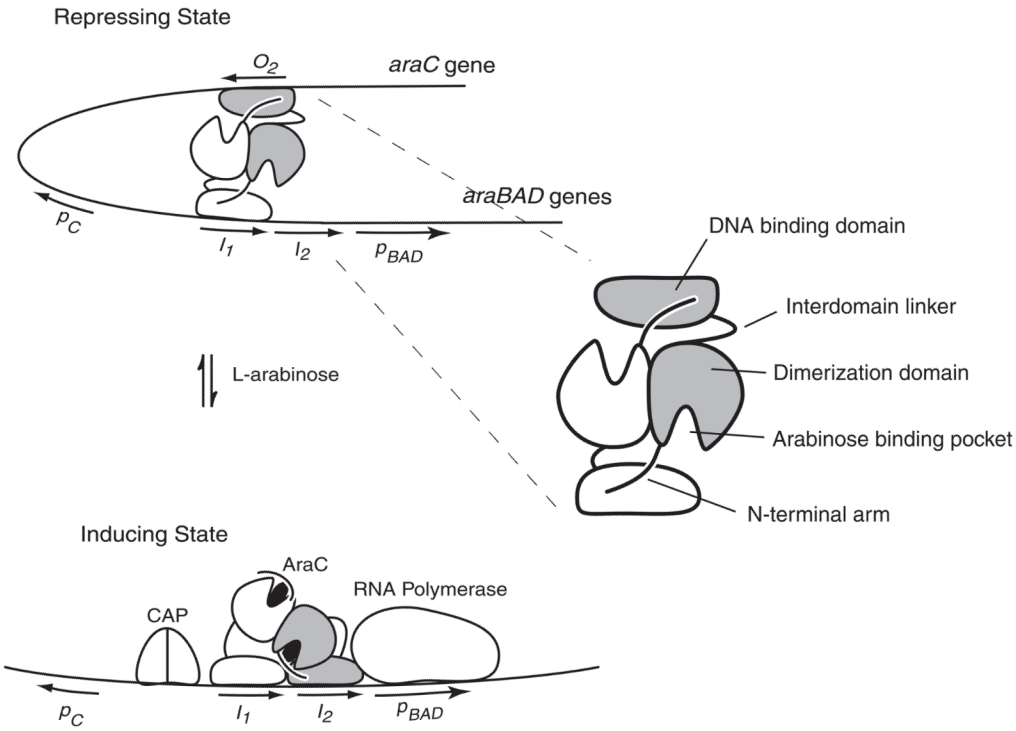

Fig. 1.

AraC and its mechanism of regulation. Schematic of the arabinose operon and AraC in its repressing and inducing states. The dimeric AraC protein in the absence of arabinose binds to $araI_1$ and $araO_2$ in a repressing state in which expression of the *araBAD* genes is low. The DNA binding and dimerization domains of AraC are connected by an interdomain linker, and the N-terminal arm is required in order that the protein bind to DNA in the repressing state. Arabinose leads the protein bind to the adjacent $araI_1$ and $araI_2$ half-sites from which it, in combination with the CAP protein, stimulate transcription of the *araBAD* genes.

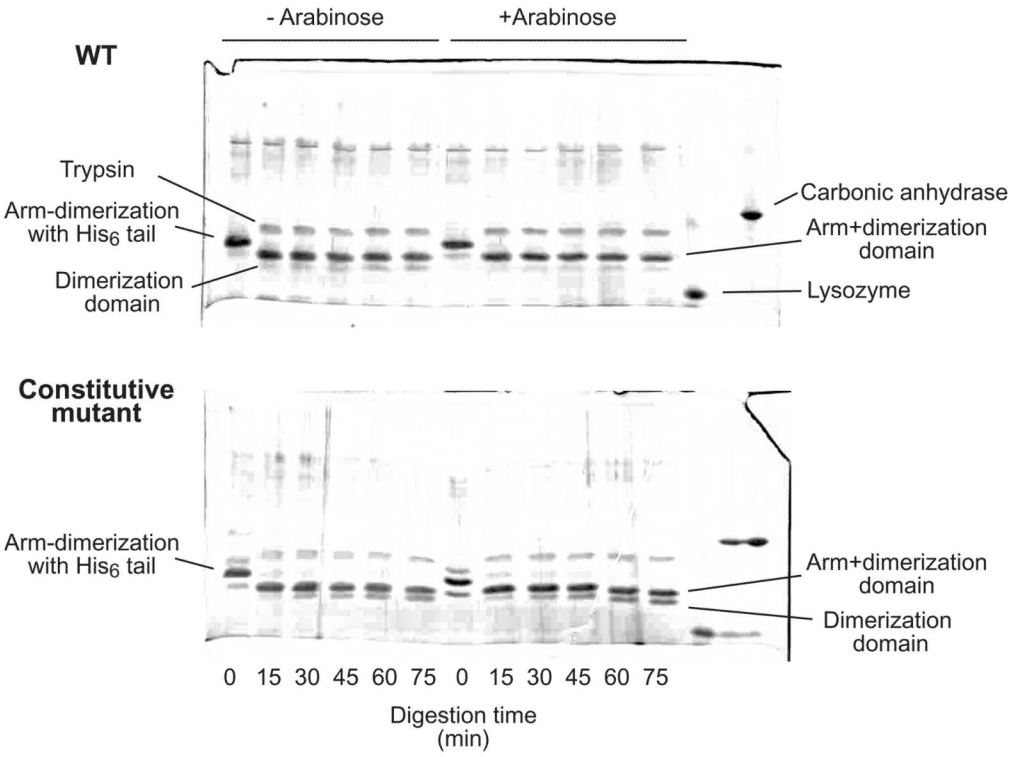

Fig. 2.

Trypsin sensitivity of arm-dimerization domain in the presence and absence of arabinose. Samples were prepared, digested, and analyzed as described in Methods. The various bands on the gel were identified by MALDI-TOF mass spectrometry. The “WT” protein carries the mutations Q5C, N16R, and C66S and the constitutive mutant protein carries in addition, L10D. Carbonic anhydrase, MW 30,000, and lysozyme, MW 14,300, were included as molecular weight standards.

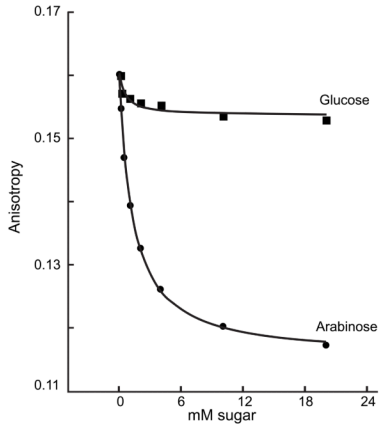

Fig. 3.

Fluorescence anisotropy of FITC-labeled AraC Q5C, N17R, C66S arm-dimerization domain in response to arabinose, circles; and glucose, squares.

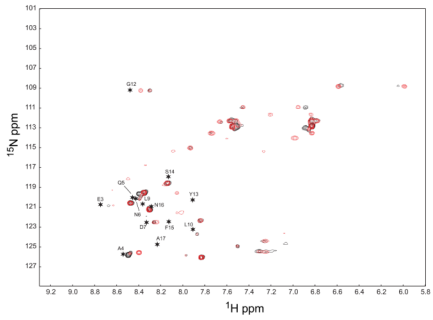

Fig. 4.

Overlay of the ^{1}H-^{15}N HSQC spectra of AraC arm-dimerization domain, residues 2–166 in the presence of arabinose (red) and in the absence of arabinose (black). The contour level threshold has been adjusted to show only the more intense peaks. The stars and labels indicate the crosspeak positions of the AraC arm residues in free arm under identical solvent conditions. No cross-peaks were observed in these positions in either spectrum. The intense peaks that occur in pairs in the ^{15}N 110–113 ppm region of the spectrum result from side chain amides.

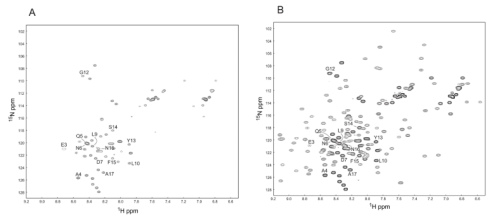

Fig. 5.

^{1}H-^{15}N HSQC spectra of the AraC Arm+22 residue linker+DBD. (A) Contour levels adjusted to reveal the intense, sharp peaks deriving from unstructured residues. The cross-peaks deriving from the backbone amides of the AraC Arm residues are labeled. (B) Contour levels set to display all the amides of the arm-DBD. The data were acquired on a Varian INOVA spectrometer operating at a ^{1}H frequency of 800 MHz and equipped with a cryogenic probe. 852 and 204 complex points were collected with sweep widths of 13333 and 3403 Hz in ^{1}H and ^{15}N resulting in acquisition times of 64ms (t_2) and 60ms (t_1), respectively. The 15N carrier frequency set at 118 ppm. Four transients were averaged for each t_1 point with a recycle delay of 1.1 second.

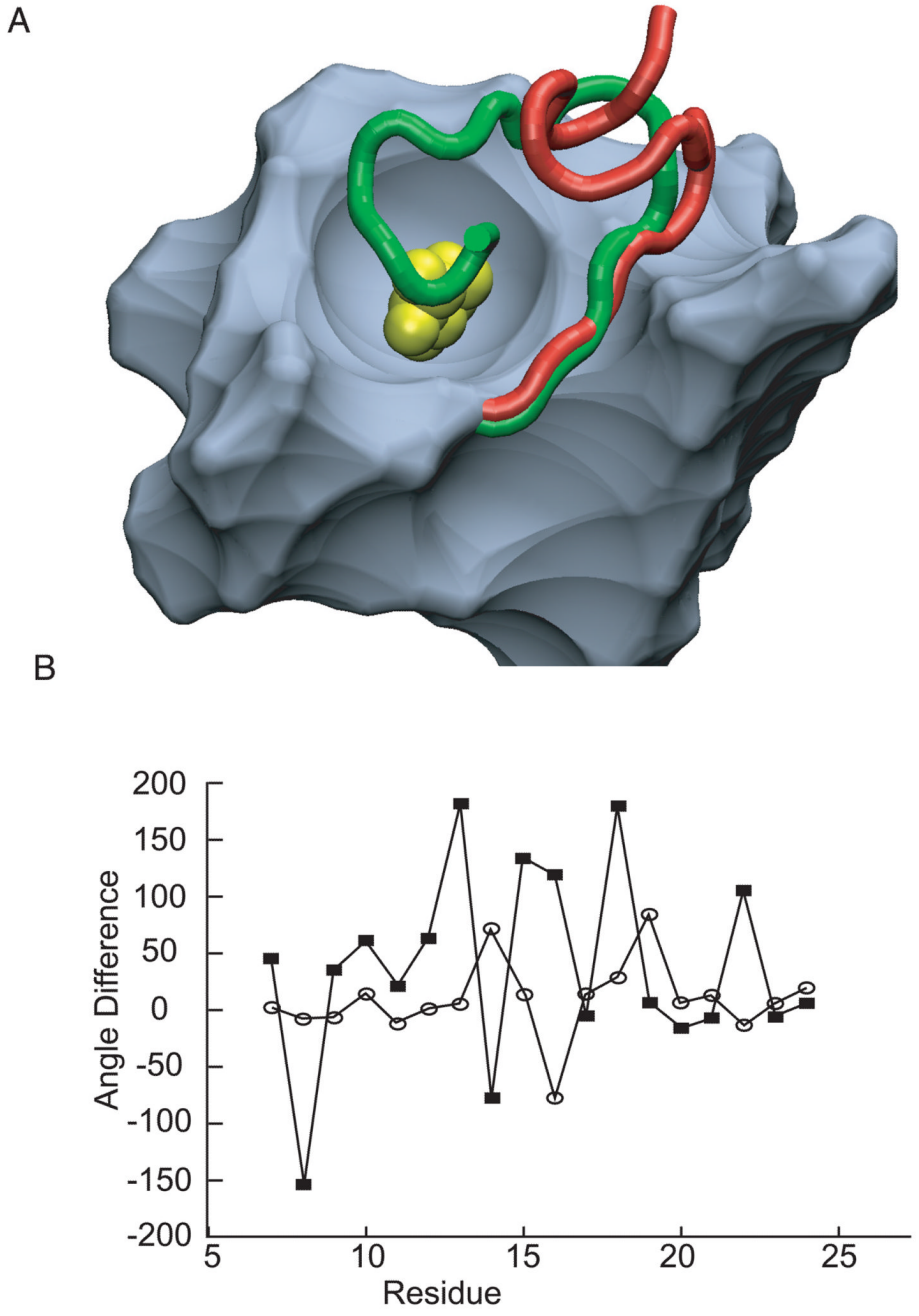

Fig. 6.

Structures of the arm and dimerization domain. (A)Tube representation of residues 7–23 of the arabinose-bound (green) and the arabinose-free state (red) structures of the dimerization domain of AraC. Residues 24–30 and 32–166 of the arabinose-bound dimerization domain, subunit A from PDB file 2ARC, were RMS overlaid with the same regions of the apo structure of the dimerization domain, subunit D, from PDB 1XJA and are shown in blue-gray. (B) Structure comparison of the N-terminal arm. Differences in the phi (open circles) and psi (filled squares) values for residues 7–25 between the arm structures in the plus arabinose (2ARC) and apo structure of the Y31V mutant (1XJA) for each position in the arm.

Table 1

Arabinose isomerase levels, units/cell, of wild type and AraC mutants

Construct	+ Arabinose	− Arabinose
WT AraC	920 ± 200	40 ± 20
AraC Q5C C66S	692 ± 140	15 ± 10
AraC Q5C N16R C66S	1300 ± 260	40 ± 20
L10D	1800 ± 360	1800 ± 360