# Atomic Force Microscopy Measurements and Model of DNA Bending Caused by Binding of AraC Protein

Mary Lowe[1], Benjamin Glezer[1], Brendan Toulan[1], and Brian Hess[1]

[1]Loyola University Maryland

April 22, 2022

**Abstract**

Atomic force microscopy (AFM) was used to conduct single-molecule imaging of protein/DNA complexes involved in the regulation of the arabinose operon of *Escherichia coli*. In the presence of arabinose, the transcription regulatory protein AraC binds to a 38 bp region consisting of the *araI1* and *araI2* half-sites. The structure of full-length AraC, when bound to DNA, was not previously known. In this study, AraC was combined with 302 bp and 560 bp DNA and arabinose, deposited on a mica substrate, and imaged with AFM in air. A statistical analysis showed that AraC induces a bend in the DNA with an angle of 69° ± 25°. By using known domain structures of AraC, geometric constraints, and contacts determined from biochemical experiments, we developed a model of the tertiary and quaternary structure of DNA-bound AraC in the presence of arabinose. The DNA bend angle predicted by the model is in agreement with the measurement value. We discuss the results in view of other regulatory proteins that cause DNA bending and formation of the open complex to initiate transcription.

## 1. Introduction

The regulatory protein AraC of the L-arabinose operon in*Escherichia coli* is a homodimer, where each subunit contains two domains: a dimerizing domain (DD) that binds arabinose (residues 1-167) and a DNA binding domain (DBD) (residues 175-281). The dimerized structure of the DD with (PDB entry 2ARC)[1] and without (PDB entry 1XJA)[2] bound arabinose has been determined by X-ray crystallography, and the DBD structure (PDB entry 2K9S) has been determined by NMR spectroscopy.[3] The DD and DBD are connected by a 7-residue interdomain linker (residues 168-174). The structure of full length AraC in the presence or absence of arabinose, when bound to DNA, is not known.

*In vivo,* under inducing conditions, that is, in the presence of arabinose, one DBD binds to the *araI1* half-site, and the other DBD binds to the nearby *araI2* with four uncontacted bases (*spacer* ) between the two.[4] Each DBD possesses two helix-turn-helix (HTH) motif regions that have been shown to contact DNA in two adjacent major grooves. When AraC is in the repressing state, it binds to the *araI1* half-site and the *araO2* half-site that is located 210 base pairs away to form a DNA loop. Upon the appearance of arabinose, the DBD that was bound to *araO2* shifts and binds to *araI2* .[5] This flexibility in binding is facilitated by the ability of the DBDs to assume a variety of orientations and positions with respect to one another.[6] Based on migration rates, DNA binding gel electrophoresis experiments indirectly indicated that in the presence of arabinose, AraC bends DNA by about 90° when the DBDs are bound to*araI1* (*I1* ) and *araI2*(*I2* ).[7]

The possible positioning of the DBDs when AraC binds to*I1-spacer-I2* are limited by the known domain structures of AraC, the length and possible conformations of the interdomain linker, residue-residue contacts and residue-base contacts determined from experiments.[6, 8, 9, 10] The length and binding energy constraints require that the DNA be appreciably bent in order to allow the two DBDs to contact the two half-sites. In this work, we determined experimentally the amount by which AraC bends DNA when it binds, and the

positioning of the DBDs to gain an understanding of the structure of DNA-bound AraC in the presence of arabinose.

Atomic force microscopy (AFM) possesses the nanometer lateral resolution that is required to measure DNA bending. Modern AFMs can examine static samples in air (sample on mica), in vacuum, under water, at low temperatures, and can measure dynamic processes at video-rate speeds.[11] AFM studies are feasible in air because during deposition onto divalent ion-treated mica substrate, DNA is able to equilibrate in two dimensions before adhering to the substrate and removal of the liquid buffer. Under such treatment, DNA is seen to resemble a worm-like chain with a persistence length of 53 nm on mica.[12] Examples of AFM studies in air include the DNA bend angle induced by the binding of protein MutS on DNA mismatches,[13] binding of Cro dimer protein to three operator sites,[14] binding of Oct-1 as a monomer to the major groove of DNA,[15] and binding of the drug cisplatin to DNA.[16] In this paper, using static AFM measurements in air, we show that binding by AraC bends DNA by 69° ± 25° (one standard deviation).

We also modeled AraC binding to *I1-spacer-I2* . Docking software was used to position each DBD onto DNA. Because no docking software is capable of handling the full flexibility of DNA,[17] we presented the software with multiple DNA conformations and chose those that yielded the best surface complementarity after computer docking. Docking software also positioned one of the DBDs onto the DD with the limitation that protein side chain flexibility was not incorporated in the process. Our goal was to determine the tertiary and quaternary structure of AraC bound to DNA. Details on hydrogen bonding and molecular dynamics to handle residue and base pair flexibility were not included in this study. Satisfactory positioning of both DBDs required a bend of the *I1-I2* binding site in the range of 52° - 88°.

## 2. Materials and Methods

### 2.1 AraC protein purification

AraC was overproduced by the pET24 expression vector (Novagen) and purified by sequential heparin and HiTrap-Q ion exchange columns, as previously described.[18] The protein stock solution contained 93 μM full-length AraC, 20 mM $Na_3PO_4$, 5 mM arabinose, 0.1 mM $NaN_3$, and 500 mM NaCl, pH 6.5.

### 2.2 DNA and solution preparation

Two DNA fragments were prepared with lengths 302 bp and 560 bp. For 560 bp DNA, target DNA containing *I1-spacer-I2* was inserted into plasmid pUC57 (Genscript). The half-sites were surrounded by 8 bp and 90 bp of wild-type sequence. Polymerase chain reaction (PCR) was used to amplify a 560-bp region containing the half-sites using primers gacggtgaaaacctctgacac (left), located within the plasmid, and ggacaactccagtgaaaagttcttc (right), located within the insert. The midpoint of *I1-spacer-I2* (38 bp in length) was located 411 bp from the left end and 149 bp from the right end, i.e., about 27% of the total DNA length from the right end (Figure 1). For 302 bp DNA,*I1-spacer-I2* was surrounded by 55 bp and 90 bp of wild-type sequence. PCR was used to amplify a 302 bp region containing the half-sites using primers cacggcagaaaagtccac (left) and tatgtagcatcaccttc (right). The midpoint of *I1-spacer-I2* was located 74 bp from the left end and 228 bp from the right end, i.e. about 25% of the total DNA length from the left end. The full sequences of the 302 bp and 560 bp DNA fragments, plasmid, insert, template DNA, and PCR conditions are shown in Figure S1.

The PCR product was spin-purified with the Qiagen PCR Purification kit and resuspended in 10 mM Tris-Cl, pH 8.5. The purity and concentration of the resulting purified DNA was determined by electrophoresis on 2% agarose gels and comparison of ethidium bromide fluorescence of the fragment to known amounts of similar sized DNA using a Low DNA Mass Ladder (Invitrogen). Purified DNA stock solution (170 nM) was used for all experiments.

DNA-only samples (Sample A) consisted of ˜65 nM DNA, ˜4 mM Tris-Cl 100 mM NaCl for 560 bp DNA (or 100 mM KCl for 302 bp DNA), 2 mM $MgCl_2$, and 10 mM HEPES. AraC/DNA samples (Sample B) contained ˜50 nM DNA, ˜3 mM Tris-Cl, 470 nM AraC, 560 μM arabinose, 89 mM NaCl (or KCl), 1.7 mM $MgCl_2$, 8.5 mM HEPES, 100 μM $Na_3PO_4$, and 0.5 μM $NaN_3$. DNA and Tris-Cl concentrations could be varied.

DNA-only and AraC/DNA samples were prepared immediately before deposition onto mica. Attachment buffer consisted of 10 mM HEPES pH7.5 and 4.03 mM $MgCl_2$.

### 2.3 AFM substrate preparation

For AFM imaging, a 10 mm diameter mica disk (grade V1, Ted Pella) was glued onto a steel disk and placed on the sample stage. The mica surface was freshly cleaved prior to sample deposition. To pre-treat the mica, 20 μl $NiCl_2$ (10 mM) were deposited on the surface,[19, 20] incubated for 1 min in a petri dish at near saturation humidity, rinsed with 2 ml distilled water, dried with flowing $N_2$ gas or air, and placed back in the humidity chamber for a few minutes. Two methods were used to bind DNA onto mica. For Method 1 ("quick mix"), 0.6 ul Sample A or B was added to 29.4 ul of attachment buffer. This combination was rapidly pipetted onto mica, incubated for 1 min, rinsed with 1-2 ml water, and dried with $N_2$ or air. For Method 2 ("injection method"), 28.6 ul of attachment buffer was first deposited onto mica and 1.4 ul of Sample A or B was added to the middle of the droplet and slightly mixed, incubated for 5 min, rinsed with water, and dried with $N_2$ or air. $MgCl_2$ was needed in the attachment buffer to improve binding of DNA. For both methods, the samples were stored in a dessicator for $> 1$ day before examination with AFM.

### 2.4 AFM imaging and data processing

AFM images were acquired in tapping mode at room temperature in air using a TT-AFM (manufactured by AFM Workshop) equipped with long cantilevers (AppNano ACLA, tip radius $< 10$ nm). In a typical experiment, 256 pixel x 256 pixel images were collected from a 1 μm x 1 μm area and scanned at a line rate of 0.7 Hz. The AFM was initially calibrated with a square grid step height standard (AppNano SHS-0.1). For work involving 560 bp DNA, a cross line grating standard was subsequently used for calibration (Ted Pella 677-AFM, line spacing 500 nm ± 1%).

All images were analyzed using the open-source software package Gwyddion, which is designed for scanning probe microscopy data visualization and analysis.[21] AFM images were processed by plane leveling, removal of background using a polynomial in the x and y directions, alignment of rows using the median, and 1D FFT filtering, as needed, to remove some of the high frequency electronic noise.

In our experiments, the DNA heights were 0.1-.45 nm above background, often resulting in low signal-to-noise. In a study that used a tip with a similar spring constant to ours, the mean height of DNA was 0.6-0.8 nm.[22] It is possible that our tip deformed the DNA. An area of exploration is to choose a cantilever with a smaller spring constant, as suggested by Asylum Research. DNA molecules could also be embedded in a layer comprised of water and salts on the mica surface leading to a short height.[22] The average surface roughness of a mica surface treated with 10 mM $NiCl_2$ is .3 nm peak-to-peak.[20]

Many articles do not state what is in the DNA and protein stock solutions used to prepare the solutions for deposition onto mica. A long period of experimentation is possibly needed to develop a suitable sample preparation protocol for first-time AFM studies of DNA and proteins. It is useful to have a readily accessible AFM for the protocol development stage.

### 2.5 Computer modeling

#### 1. Starting structures

To model the AraC/DNA complex we used the structure of the dimerized, arabinose-bound DD determined by X-ray crystallography (2ARC) and the structure of the DBD determined by NMR (2K9S). Residues 7-167 are in the structure of both DDs along with a portion of the linker, residues 168-170. The full linker, residues 168-174 (NESLHPP), is present in the DD, chain C, with no L-arabinose (1XJA).

#### 2. Positioning the DBDs onto DNA

*Overview and biochemical contacts*

The first step in building a full model of AraC was to position each DBD onto a half-site using known residue-base interactions determined biochemically and geometric analysis of how much the DNA half-site needed

to be bent in order to achieve satisfactory binding to the DBD. We refer to the two AraC DBDs as DBD-A and DBD-B. To approximately position the DBDs on the *I1-I2* half-sites, we used the HDOCK webserver[17] to initially place DBD-A onto *I1* and DBD-B onto *I2* . HDOCK uses a rigid-body docking approach and can incorporate biochemical contacts to constrain the number of putative binding modes. S209 and H213 in the first HTH of the DBD were specified as these residues are known to contact the major groove of DNA at bases -69 and -48 (modern residue numbering).[8, 10] Q258 also makes a weak contact. The top 100 models can be downloaded.

*Criteria for docking*

We had four criteria for acceptable docking. (1) DBD-A and DBD-B both had to appear to make protein-DNA contacts that have been demonstrated biochemically. The first HTH makes stronger contacts with the DNA[8] than the second HTH and had to sit properly in the major groove. As best as possible, the second HTH also needed to be over the major groove. (2) Each half-site had to be roughly centered on the appropriate DBD. (3) Using the surface representation feature in PyMol (https://pymol.org), the docked structures had to have good surface complementarity but could not intersect each other. (4) Because AraC has been shown experimentally to bind to direct-repeat*I1-I2* ,[6] we chose docked structures with the DBDs oriented as indicated in the biochemical contact experiments in which the second HTH is closer to RNA polymerase (see Discussion).

*DNA segment method*

To achieve plausible docking, it was necessary for us to make two types of conformation changes to the DNA before presenting it for docking. The first type, called "DNA segment method," involved making adjustments at the base-pair level by using experimentally-determined structures for DNA segments about 4-5 bp in length. The *I1* sequence was divided into six overlapping segments, for example, TAGC, GCATT, TTTTT, etc. For each segment, a search for the sequence was performed in the RCSB database and at least four examples were chosen. The geometric parameters of the sequence were calculated using Web 3DNA (see below). After averaging the parameter values for the desired region of the DNA over the four example sequences, a table was constructed with the geometric parameters of the entire *I1* sequence. Web 3DNA was then used to rebuild the *I1* structure from the geometric parameters. The same procedure was followed for *I2* and the spacer between*I1* and *I2* . Over 52 experimental structures for DNA were used to build the *I1-spacer-I2* sequence.

*Large-scale bending*

Using the DNA-segment method only, initial attempts at docking the DBDs onto *I1* or *I2* were not satisfactory. A second type of DNA conformation change was applied to conduct large-scale bending of DNA using Web 3DNA 2.0, a web-based server[23] designed to manipulate DNA conformations. The software generates rigid-body values (three translations and three rotations) of a single base pair (bp) and six step parameters for neighboring bps. It allows modification of the bp and step parameters to build customized DNA structures.

To calculate the bp step parameters to bend DNA within a plane, we followed the recommendations in Calladine *et al* .[24] DNA roll angles were modified gradually at each bp step in the relevant region in which the roll angle $R_n$ at step n is given by , where a

larger A results in a greater total bend angle and b introduces a phase angle to position the maximum roll angle at the desired location along the DNA. This formula makes the approximation that a full helical twist of B-DNA contains 10 bp/turn and that the roll angle has the same period of 10 steps (also see Discussion). Values for A and b were selected empirically by presenting HDOCK with a variety of bent DNA structures and choosing the docked structure with the best surface complementarity.

*DBD-A/I1 and DBD-B/I2 complexes*

After applying the DNA segment method, we placed high priority on the known biochemical contacts and

made sure that the first HTH of *I1* docked satisfactorily onto DBD-A. This resulted in a large gap between the second HTH and DNA (Figure 2A). Since *I1* contains an A-tract with a length of 5 bp, we knew that a bend had to occur in the neighborhood of 11-28deg.[25] Large-scale bending of DNA was then performed starting within the A-tract, and a more successful dock resulted (Figure 2B). For the DBD-B/*I2* complex (Figure 2C), less bending was needed than for DBD-A/*I1* . Values for roll adjustments to achieve bending are given in Table 1.

## 3. Results

We used AFM to measure the DNA bend angle when DNA is bound to AraC in the presence of arabinose. We also created a model for the AraC/DNA complex by incorporating biochemical contacts, positioning the domains using docking software, applying geometric constraints, and bending DNA. The model prediction and AFM measurements agree.

### 3.1 AFM results

DNA consisting of *I1-spacer-I2* was deposited on a mica substrate and imaged with AFM. For 302 bp DNA, a comparison is shown for a DNA-only sample (Figure 3A) and samples prepared with DNA, AraC, and arabinose (Figures 3B, C). For DNA-only, most structures have a curvilinear shape. When AraC was added, DNA exhibited a sharper bend in many structures, where, in most of the cases, the bend was occurring about one-third of the length from one end, consistent with the asymmetric placement of *I1-spacer-I2* within the 302 bp sequence. For simple structures, the interior angle $\varphi$ of the bend was measured by drawing tangents to the contour on both sides of the bend, as shown in Figure 4A. Because the 302 bp DNA was so short, the tangents were effectively constructed near the ends of the DNA. The DNA bend angle $\vartheta$ was calculated from $\vartheta = 180^\circ - \varphi$. For each $\vartheta$ histogram, a Gaussian function was fit to all nonzero $\vartheta$. The average bend angle and standard deviation for 302 bp DNA-only samples was $47^\circ \pm 27^\circ$ (Figure 5B). When AraC and arabinose were added, $\vartheta = 75^\circ \pm 26^\circ$ (Figure 5C); DNA bending was stronger.

DNA bend angles were also measured for 560 bp DNA. In comparison with 302 bp DNA (Figure 3A), 560 bp DNA contained more wiggles (Figure 3D), and it was not possible to determine exactly where to measure a bend angle. However, when AraC and arabinose were added to 560 bp DNA, many structures showed a pronounced checkmark shape (Figures 3E, F). A histogram was constructed for simple structures resulting in $\vartheta = 65^\circ \pm 21^\circ$ (Figure 5D). This average for 560 bp AraC/DNA complexes was somewhat less than for 302 bp AraC/DNA, suggesting that the longer length may assist with reducing the bend angle when DNA is deposited on a mica substrate. Because the $\vartheta$ values for 302 bp and 560 bp overlapped significantly, we combined the bend angle data for both (Figure 5A) and obtained $\vartheta = 69^\circ \pm 25^\circ$.

Many types of DNA structures appeared on the mica substrates. The circular features in Figs. 3B, C, E, F represent AraC, AraC/DNA complexes, or aggregates. For AraC/DNA samples, the heights of the circular features were .12-.47 nm, suggesting that the shorter features were protein while the taller ones were aggregates. For DNA, criteria had to be established on what to measure. We decided to not measure U- and J-shaped structures (Figure 4B) as well as segmented structures (Figure 4C). In many cases, there was a small hook at the end of the DNA (Figure 4A) or a small circular feature (Figure 4C); these were ignored. Shorter DNA fragments also appeared perhaps due to impurities or incomplete synthesis in the PCR reactions. For 302 bp and 560 bp DNA, we did not measure structures shorter in length than 80 nm and 135 nm, respectively.

The question arises whether the protein is visible in the AraC/DNA structure. Due to AFM tip broadening effects and noise in the images, we cannot conclusively state whether the protein is visible. Shorter circular features on the mica, which may be protein, are in the same height range as most of the DNA. Based on the model (see below), we would expect that the protein would not lie on top of the DNA; instead it would be near the corner of a checkmark next to the DNA.

For 560 bp DNA, the average length for DNA samples was $159 \pm 24$ nm (Figure S2), shorter than what would be expected in the presence of $NiCl_2$ and $MgCl_2$ (see Discussion). Uncertainties in the length measurements

could arise from AFM tip broadening and imprecise contour determination.

### 3.2 Constructing the model of full-length AraC and DNA in the presence of arabinose

In initial attempts to build a model of full-length AraC bound to DNA in the presence of arabinose, we started with a straight helical structure for *I1* and *I2* generated by a computer. However, we were not able to achieve a satisfactory dock of straight *I1* onto the DBD. We then developed a DNA segment method (see Methods) to build*I1* and *I2* out of experimentally-determined positions of DNA atoms from the RCSB database. Upon docking, this process resulted in an acceptable positioning of one of the HTHs of the DBD onto a half-site. To dock both HTHs of each DBD onto a half-site, we had to bend the DNA by adjusting the DNA roll values (Figures 2B, C).

Mutagenesis experiments have shown that there are points of contact between N194 of the DBD and H126 and N139 of the dimerization domains (DDs).[9] H126 and N139 lie on the DD of different subunits and are close to each other. Using this information and HDOCK software, the DBD-A/*I1* complex was positioned onto the DDs with N194 of DBD-A close to N139 of DD B and H126 of DD A (Figure 6). Due to linker length constraints, the N-terminus of DBD-A had to be near the C-terminus of DD A. (DD A and linker A are parts of subunit A.) The distance between $C_\alpha$ of I167 of DD A and $C_\alpha$ of M175 of DBD-A is 23 Å, which is consistent with the length of linker A.

Biochemical measurements have shown that the two DBDs are bound to*I1-I2* with a direct repeat orientation.[6] To build the model, the DBD-B/*I2* complex had to be brought closer to DD B, as shown in Figure 6. Large-scale bending of DNA was accomplished in the 4-bp spacer region and first two bp of *I2* . This process enabled $C_\alpha$ of M175 of DBD-B to be within 25 Å of $C_\alpha$ of I167 of DD B (Figure 7). Linkers A and B are uncoiled relative to the conformation appearing in the crystal structure 1XJA. Table 1 shows the values used in adjusting the roll angles to conduct large-scale bending of the DNA in the extended spacer region.

The full model of AraC/DNA is shown in Figure 7A. The two orange line segments indicate a bend angle of 69°. Different lines can be constructed tangent to the *I1* contour resulting in a range of possible DNA bend angles from 52° to 88°. Two straight line segments (teal) indicate where the linkers are located. Linker B is clearly visible in the middle of the figure; linker A is near the top and is rotated to be almost hidden from view. Figure 7B and the schematic diagram Figure 7D show the large-scale bend in the DNA and the domains of AraC. The DDs are oriented such that DD B is on top of DD A. Figures 7C, E represent a rotated view of AraC/DNA in which the DNA appears almost straight and both DDs are visible. The 5' end of *I1* is near linker A and the N-terminus of DBD-A. Similarly the 5' end of*I2* is near the N-terminus of DBD-B and linker B.

A Pymol session file (.pse) of the AraC/DNA complex is included in the Supporting Information (Figure S3). Explicit hydrogens are not present in the structure because HDOCK and Web 3DNA remove them. In the HDOCK scoring function, the effects of hydrogens are incorporated in the potentials associated with different heavy atom types and their environments.[26]

## 4. Discussion

*DNA bending caused by the binding of AraC to DNA*

In this work we studied the structure of the AraC/DNA complex using AFM to measure the DNA bend angle caused by the binding of AraC to the*araI1 -araI2* region of DNA. AFM is able to produce a direct view of bending at the single molecule level without the need for labeling the DNA or protein. Independently of the measurements, we built a model using known AraC domain structures, biochemical contacts between AraC and DNA, and modifications of the conformation of DNA in order to be consistent with all dimensions. The measured DNA bend angle and the one predicted in the model agree within the uncertainties. We believe that the model shows the likely overall positions of the basic components (DDs, DBDs, DNA) of AraC bound to DNA.

The measured bend angle distribution for AraC/DNA complexes has a peak at 69° ± 25°. A possible reason for the broad distribution is that the 7-residue linker of the DBD bound to *I2* is flexible, leading to a range of angles. Experimental evidence indicates that the interdomain linker is highly flexible and most likely not an alpha helix.[6] *In vivo* , AraC binds tightly to the two direct repeat half-site sequences *araI1-araI2.* Experiments also show that AraC can bind tightly to inverted (*araI1 -invaraI1* ) DNA half-sites, indicating the ability of the linker to change its coiling conformation. Another possible reason for the broad distribution is that the AraC/DNA complexes would land on the mica substrate in different ways; a collapse of DNA onto mica could appear in the AFM scan to have a range of bend angles.

All of our bend angle distributions have a peak at 0°. There could be unbound DNA on the mica surface. Using a dissociation constant $K_d$ = 2 x $10^{-12}$ M for AraC – DNA binding in the presence of arabinose,[27] calculations show that prior to deposition on mica, all of the DNA should be bound to AraC. However in the final deposition and washing steps, it is possible that some of the DNAs became unbound. In addition, AraC/DNA complexes might have landed on the surface in an orientation where the DNA would appear straight (Figure 7C). If that were the case, one would expect that the structure would have a higher height where the DNA and protein overlap. We did not observe this effect and cannot make a definitive statement due to the presence of aggregates and the noise level in the images.

We can speculate on the role of DNA bending. AraC is a transcription regulatory protein, and like other DNA binding proteins, AraC causes DNA to bend. Perez-Martin and Espinosa[28] proposed that increases in DNA curvature would facilitate RNA polymerase-promoter interactions during the initiation of transcription. In the arabinose operon in *E. coli* , upon the addition of arabinose, AraC binds to*araI1* and *araI2* , and AraC and cyclic AMP receptor protein (CRP) both help RNA polymerase (RNAP) bind to the $p_{BAD}$promoter as well as speed the formation of open complex to initiate transcription.[29, 30] We conjecture that substantial bending of the DNA binding sites of CRP, AraC, and RNAP would allow a compact structure to form that could involve specific contacts among the three proteins and improve protein-DNA interactions.[15]Bending of DNA may also assist melting of the double-stranded structure in the region where transcription is to begin, like sharp bending of a rope separates its strands.

A schematic diagram is shown in Figure 8 of CRP, AraC, RNAP and DNA. The dashed lines show a potential path of DNA where CRP has induced a DNA bend angle of 80°, AraC a bend of 69°, and RNAP a bend in between 30°-90°, as suggested by experimental data.[31, 32, 33]As can be seen, somewhat greater bending anywhere in the complex could bring CRP and RNAP into contact. The bending of DNA by AraC appears to facilitate the process of forming a compact structure.

*Other experiments where DNA bend angle was measured*

The bend angle distribution, 69° ± 25°, measured in this work is relatively broad. The λ Cro transcription regulatory protein produces DNA bending of 69° ± 11°.[14] The cellular transcription factor Oct-1 produces a 42° ± 12° DNA bend angle.[15]Broad distributions with a standard deviation in the neighborhood of our work were found when the DNA repair protein MutS binds to homoduplex and heteroduplex DNA.[13] In a study involving the binding of the drug cisplatin to DNA,[16] simulations of 300 bp DNA were conducted with an adjustable bend angle. AFM-like images of simulated DNA molecules show that DNA mostly has a curvilinear shape for 0° induced bend angles, as we saw in our experiments. When a bend angle is introduced, a range of angles is seen due to thermal and mechanical effects. For a persistence length of 50 nm and a bend angle of 60°, 2-3% of the structures are straight. In our experiment, we found a higher percentage of straight structures suggesting that the protein became unbound from the DNA due to sample preparation methods.

*Limitations of computer modeling*

To produce the large-scale bends in DNA, we adjusted the roll angles using an approximation of 10 bp/turn. In experiments, biochemical analyses have shown that straight DNA in solution has 10.4 ± 0.1 bp/turn, rather than 10 bp/turn as observed in the solid state.[34] *In vivo* , supercoiling of DNA leads to a slight unwinding with the result of 11.1 bp/turn.[35]Because the corresponding parts of *I1* and *I2* half sites are 21 bp apart, it suggests that in this region of DNA, the twist is closer to 10.5 bp/turn to have the DBDs

bind on the same side of the DNA. For 10, 10.4, and 11.2 bp/turn, the total twist over the 38 bp length of *I1-spacer-I2* is three full helical turns and 288°, 235°, and 152° of the fourth helical turn, respectively. The incomplete fourth turn would affect which side of the DNA the DBDs could bind to and the relative rotation of the two DBDs. For producing tertiary and quaternary models of protein/DNA complexes, the value of 10 bp/turn was adequate for bending the DNA in a narrow step range (< 10 steps) but it introduces inaccuracies in the hydrogen bonding within DNA, the plane of the bend, and the relative orientation of the two DBDs.

To create the full model of the AraC/DNA complex, docking software was used as one stage of the process to form protein-DNA and protein-protein complexes. Our docking procedure however, cannot produce reliable results for residue/base, base/base, and residue/residue interactions. Some docking packages support the DNA sequence as input.[17] However, we found that when the DNA structure is generated by a computer from a sequence, the DBD could not be satisfactorily docked onto DNA. We had to divide the sequence into segments and construct each segment from experimentally-determined coordinates. While the DNA segment method initially positioned the DBD on the DNA, it was still insufficient to achieve satisfactory docking in that both HTHs of the DBD did not sit fully in the major groove of DNA. Experiments have shown that protein and DNA both experience conformational changes upon binding, and that when protein bends DNA, there is an energy cost associated with distortion such that the DNA is not in its lowest energy configuration.[14] Docking software should ideally allow for DNA flexibility and protein backbone and side chain flexibility, which are difficult features for any software package to possess.[36, 37] We addressed this issue partially by manually adjusting the bend in the DNA and submitting the 3D structure to the docking software. This process was repeated until satisfactory surface complementarity was achieved. However detailed manipulations on the atomic scale were not conducted that would result in the true energy of the DNA or the protein. We also observed that upon docking, the two HTHs of the DBD can sit in the DNA major groove in various ways. We had to make a judgment on which model was the best; inaccuracies could have been introduced at this stage. Finally the docking approach in this paper did not take into account effects of solvation and salt.[39] As a step towards determining tertiary and quaternary structure of the AraC/DNA complex, docking was useful; however conclusions about structure and interactions on an atomic scale must be drawn carefully.

*DNA contour length*

Based on X-ray crystallography, B-form DNA has a helical rise of 3.38 Å/bp, and therefore our 560 bp DNA would be expected to have a contour length of 189 nm instead of the 159 ± 24 nm length that we measured in the absence of protein. This result was doublechecked by using an AFM calibration standard with a 500 nm ± 1% period, comparable to the image size. Possible physical reasons for the shortening could be associated with the immobilization of DNA on the mica substrate. In the presence of $NiCl_2$ and Tris, the average helical rise is 2.97 ± 5% Å/bp.[39] Shortening of 500 bp DNA by about 10 nm occurs in the presence of $Mg^{2+}$ or $Mn^{2+}$ due to partial conformational transition from B-DNA to A-DNA upon drying.[40] The drying process, phosphate-cation interactions on the mica surface, and cations that bridge two parts of the DNA backbone[40] may contribute to shortening.

## 5. Conclusions

Traditional methods of determining structure have enabled the DD and DBD structures to be determined by X-ray crystallography and NMR spectroscopy, respectively. These techniques, however, could not determine the entire structure of AraC nor the structure of AraC bound to DNA. AFM is capable of imaging bending of DNA for an ensemble of DNA conformations when small proteins such as AraC (63 kDa) are bound to DNA. The bend angle measurements can be used to validate models constructed from biochemical information, geometric constraints, software for rebuilding DNA from bp and step parameters, and positioning using docking software.

## Acknowledgments

We wish to thank R. Schleif for comments and suggestions throughout the course of this work. We also thank A. Kutt, P. Kutt and personnel at AFM Workshop. This work was supported by funds from Loyola

University Maryland.

**Supporting Information**

Supporting information is available with the electronic version of this article. It includes (1) full sequences of the plasmid, insert, primers, and PCR conditions for producing the 302 bp and 560 bp DNA fragments used in this study, (2) histogram of 560 bp DNA lengths, and (3) PyMol session file of AraC bound to DNA and arabinose

**References**

1. Soisson SM, MacDougall-Shackleton B, Schleif R, Wolberger C. Structural basis for ligand-regulated oligomerization of AraC.*Science* . 1997;276:421–425. DOI:10.1126/science.276.5311.421.

2. Weldon JE, Rodgers ME, Larkin C, Schleif RF. Structure and Properties of a Truly Apo Form of AraC Dimerization Domain. *PROTEINS: Structure, Function, and Bioinformatics* 2007;66:646–654.

3. Rodgers ME, Schleif R. Solution structure of the DNA binding domain of AraC protein.*Proteins* 2009;77: 202-208.

4. Schleif R. AraC protein, regulation of the L-arabinose operon in*Escherichia coli* and the light switch mechanism of AraC action.*FEMS* 2010;34:779-796. DOI:10.1111/j.1574-6976.2010.00226.x

5. Lobell R and Schleif RF. DNA Looping and Unlooping by AraC Protein.*Science* 1990;250:528-532.

6. Carra JH, Schleif RF. Variation of half-site organization and DNA looping by AraC protein. *The EMBO Journal* 1993;12(1):35-44.

7. Saviola B, Seabold RR, Schleif RF. DNA Bending by AraC: a Negative Mutant. *J. Bacteriology* 1998;180 (16):4227-4232.

8. Brunelle A, Schleif R. Determining residue-base interactions between AraC protein and *araI* DNA. *J. Mol. Biol.* 1989;209:607-622.

9. Schleif R. 2021. A Career's Work, the L-Arabinose Operon: How it functions and how we learned it. *EcoSal Plus* 2021. https://doi.org/10.1128/ecosalplus.ESP-0012-2021

10. Greenfield L, Boone T, Wilcox G. DNA sequence of the araBAD promoter in *Escherichia coli* B/r. *Proc. Nati. Acad. Sci. USA*1978;75(10):4724-4728.

11. Main KHS, Provan JI, Haynes PJ, Wells G, Hartley JA, Pyne ALP. Atomic force microscopy—A tool for structural and translational DNA research. *APL Bioeng.* 2021:5:031504. DOI:10.1063/5.0054294

12. Rivetti C, Guthold M, Bustamante C. Scanning Force Microscopy of DNA Deposited onto Mica: Equilibration versus Kinetic Trapping Studied by Statistical Polymer Chain Analysis. *J. Mol. Biol.*1996;264:919–932.

13. Tessmer I, Yang Y, Zhai J, Du C, Hsieh P, Hingorani MM, Erie DA. Mechanism of MutS Searching for DNA Mismatches and Signaling Repair.*J. Biological Chemistry* 2008;283(52):36646–36654.

14. Erie, DA, Yang G, Schultz HC, Bustamante C. DNA Bending by Cro Protein in Specific and Nonspecific Complexes: Implications for Protein Site Recognition and Specificity. *Science* 1994;266:1562-1566.

15. Mysiak, ME, Wyman C, Holthuizen PE, van der Vliet PC. NFI and Oct-1 bend the Ad5 origin in the same direction leading to optimal DNA replication. *Nucl. Acids Res.* 2004;32:6218-6225.

16. Dutta S, Rivetti C, Gassman NR, Young CG, Jones BT, Scarpinato K, Guthold M. Analysis of single, cisplatin-induced DNA bends by atomic force microscopy and simulations. *Journal of Molecular Recognition* 2018; 31(10):e32731, 11 pages total, DOI: 10.1002/jmr.2731

17. Yan Y, Tao H, He J, Huang S-Y. The HDOCK server for integrated protein–protein docking. *Nat. Protoc.* 2020;15:1829-1852. DOI:10.1038/s41596-020-0312-x

18. Rodgers ME, Schleif R. Heterodimers reveal that two arabinose molecules are required for the normal arabinose response of AraC.*Biochemistry* 2012;51:8085-8091.

19. Pastre D, Pietrement O, Fusil S, Landousy F, Jeusset J, David MO, Hamon L, Le Cam E, Zozime A. Adsorption of DNA to Mica Mediated by Divalent Counterions: A Theoretical and Experimental Study.*Biophysical Journal* 2003;85:2507–2518.

20. Hsueh C, Chen H, Gimzewski JK, Reed J, Abdel-Fattah TM. Localized Nanoscopic Surface Measurements of Nickel-Modified Mica for Single Molecule DNA Sequence Sampling. *ACS Appl. Mater. Interfaces*2010;2(11): 3249–3256. DOI:10.1021/am100697z

21. Nečas D, Klapetek P. Gwyddion: an open-source software for SPM data analysis. *Cent. Eur. J. Phys.* 2012;10(1):181-188. Also see http://gwyddion.net/ .

22. Margeat E, Le Grimellec C, Royer CA. Visualization of trp Repressor and its Complexes with DNA by Atomic Force Microscopy. *Biophysical Journal* 1998;75:2712-2720.

23. Li S, Olson WK, Lu XJ. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. Nucleic Acids Research 2019;47:W26-W34. DOI: 10.1093/nar/gkz394

24. Calladine CR, Drew H, Luisi B, Travers A. Understanding DNA: The Molecule and How it Works. New York: Elsevier Academic Press; 2004. 72-78 p.

25. Stefl R, Wu H, Ravindranathan S, Sklenář V, Feigon J. DNA A-tract bending in three dimensions: Solving the $dA_4T_4$ vs. $dT_4A_4$ conundrum. *Proc. Natl. Acad. Sci.* 2004;101(5):1177-1182. DOI:10.1073/pnas.0308143100

26. Huang S-Y, Zou X. A knowledge-based scoring function for protein-RNA interactions derived from a statistical mechanics-based iterative method. *Nucleic Acids Research* 2014;42(7):e55. DOI:10.1093/nar/gku077

27. Hendrickson W, Schleif RF. Regulation of the Escherichiu coli L-Arabinose Operon Studied by Gel Electrophoresis DNA Binding Assay.*J. Mol. Biol.* 1984;174:611-628.

28. Perez-Martin J, Espinosa M. Protein-induced Bending as a Transcriptional Switch. *Science* 1993;260(5109):805-807.

29. Zhang X, Reeder T, Schleif R. Transcription Activation Parameters at*ara* $p_{BAD}$ . *J. Mol. Biol.* 1996;258:14–24.

30. Zhang X, Schleif R. Catabolite gene activator protein mutations affecting activity of the *araBAD* Promoter. *J Bacteriology*1998;180:195-200.

31. Benoff B, Yang H, Lawson CL, Parkinson G, Liu J, Blatter E, Ebright YW, Berman HM, Ebright RH. Structural Basis of Transcription Activation: The CAP-αCTD-DNA Complex. *Science* 2002;297(5586):1562-1566.

32. Glyde R, Ye F, Jovanovic M, Kotta-Loizou I, Buck M, Zhang X. Structures of Bacterial RNA Polymerase Complexes Reveal the Mechanism of DNA Loading and Transcription Initiation. *Molecular Cell* 2018;70: 1111–1120. DOI:10.1016/j.molcel.2018.05.021

33. Lee J, Borukhov S. Bacterial RNA Polymerase-DNA Interaction-The Driving Force of Gene Expression and the Target for Drug Action.*Front. Mol. Biosci.* , 2016;3:73. DOI: 10.3389/fmolb.2016.00073

34. Wang JC. Helical repeat of DNA in solution. *Proc. Natl. Acad. Sci.* 1979;76(1):200-203.

35. Lee D, Schleif R. In vivo DNA Loops in *araCBAD* : Size Limit and Helical Repeat. Proc. Nat. Acad. Sci. 1989;86:476-480.

36. He J, Tao H, Huang SY. Protein-ensemble–RNA docking by efficient consideration of protein flexibility through homology models.*Bioinformatics* 2019*;* 35(23):4994–5002.

37. Pinzi L, Rastelli G. Molecular Docking: Shifting Paradigms in Drug Discovery. Int. J. Mol. Sci. 2019;20: 4331. DOI:10.3390/ijms20184331

38. Pagadala NS, Syed K, Tuszynski J. oftware for molecular docking: a review. *Biophys Rev* 2017;9:91–102. DOI 10.1007/s12551-016-0247-1

39. Sanchez-Sevilla A, Thimonier J, Marilley M, Rocca-Serra J, Barbet J. Accuracy of AFM measurements of the contour length of DNA fragments adsorbed on mica in air and in aqueous buffer. *Ultramicroscopy*2002;92:151–158.

40. Japaridze, A., Vobornik D, Lipiec E, Cerreta A, Szczerbinski J, Zenobi R, Dietler G. Toward an Effective Control of DNA's Submolecular Conformation on a Surface. Macromolecules 2016;49:643-652. DOI: 10.1021/acs.macromol.5b01827

**List of Tables and Figures**

Table 1. Parameters used to adjust the roll angles to bend the DNA: . Base pair numbering: *I1* starts at n = 1; *I2* starts at n = 1; spacer starts at n = 18. Sequence "spacer-ext" consists of the 4-bp spacer region between *I1* and *I2* and the first two bp of*I2* .

Figure 1. 560-bp amplicon comprised of the *araI1* and *araI2*half-sites. The black region contains the sequence for *I1 – spacer – I2* : *tagcatttttatccata* agat*tagcggatcctacctga* in which the half-sites are underlined. *I1* is on the left. The white region precedes *I1-spacer-I2* and has a length of 392 bp. The striped region has a length of 130 bp. The 302-bp amplicon is similar except that *I1-spacer-I2* is located closer to the left end. Refer to Figure S1 for the full sequence information.

Figure 2. Surface representations of DBD-DNA docking. N-terminus of DBD (white) and 5' end of DNA (gray) are shown by the black regions. (A) Initial docking of DBD/*I1* . A large gap exists. (B) DBD/*I1*complex after bending the DNA to reduce the gap. (C) DBD/*I2* after slight bending of DNA.

Figure 3. AFM images of samples prepared with DNA only and samples with DNA, AraC, and arabinose. More sharply bent DNA is visible in samples prepared with AraC and arabinose. The circular features contain protein. Width of image = 500 nm. (A) 302 bp DNA only, (B and C) 302 bp DNA+AraC+arabinose. (D) 560 bp DNA only. Bend angles are difficult to measure. (E-F) 560 bp DNA+AraC+arabinose.

Figure 4. AFM images showing different DNA structures on mica for 560 bp DNA. Width of image = 250 nm. (A) Simple structure showing how bend angles were measured. The red lines are tangent to the contour on either side of the bend. Bend angles were not measured for (B) highly curved and (C) segmented structures. A small hook at the end of the DNA, shown in (A), or a circular feature, shown in (B) and (C) was ignored for bend angle measurements.

Figure 5. DNA bend angle histograms. Gaussian functions were fit to the data, excluding 0° bend angles. Mean and standard deviation are shown. (A) Samples prepared with DNA, AraC, and arabinose. Data for 302 bp and 560 bp were combined. Average bend angle: 69° ± 25°, N = 267; (B) 302 bp DNA only, 47° ± 27°, N = 88; (C) 302 bp DNA, AraC, arabinose, 75° ± 26°, N = 110; (D) 560 bp DNA, AraC, arabinose, 65° ± 21°, N = 157.

Figure 6. Initial construction of the AraC/DNA model prior to large-scale bending of DNA in the spacer region. (A) DBD-A/*I1*complex is docked onto both DDs. *I1* (medium gray); DBD-A (white); DDs (dark gray). Surface representation. (B) Full *I1-spacer-I2* is shown with the DBD-A/*I1* /DDs complex (left) and DBD-B/*I2*complex (right). Cartoon representation.

Figure 7. Full model showing the bending of DNA. N = N-terminus. (A) Cartoon representation. Orange line segments show a bend angle of 69°. DDs (dark gray), DBD-A (light blue), DBD-B (dark blue), DNA (gray), 4-bp spacer region (pink). *I1* is above the spacer. Teal lines, representing the 7-residue linkers, are drawn

between $C_\alpha$of I167 of the DD and M175 of the DBD. (B) Surface representation showing DNA bending. (C) Alternate view rotated by 90° from (B). Both DDs are visible. (D) Schematic diagram corresponding to (B) indicating the domains of AraC and *I1-spacer-I2* . (E) Schematic diagram corresponding to (C).

Figure 8. A schematic diagram approximately to scale of the sizes, relative binding positions, and DNA bends induced by CRP, AraC, and RNAP.

Figure S1. Full sequences of the plasmid, insert, primers, and PCR conditions for producing the 302 bp and 560 bp DNA fragments used in this study.

Figure S2. Histogram of measured lengths of 560 bp DNA. No AraC or arabinose was added. A Gaussian function was used to fit the histogram data and determine the mean and standard deviation of the length distribution. Length = 159 ± 24 nm, N = 131.

Figure S3. PyMol session file showing the model of AraC and bent DNA. The user can click on the DD and DBD domains of subunits A and B;*araI1* , *araI2* , spacer, key biochemical contacts in the protein and DNA, and the $C_\alpha$ endpoints of linkers A and B. Filename: Figure_S3_structure_AraC_DNA.pse.

| **Sequence** | **n (bp)** | **A (°)** | **b (bp)** |
|---|---|---|---|
| *I1* | 9 - 14 | 16.0 | 5.5 |
| *I1* | 15 - 17 | 12.0 | 10.5 |
| *I2* | 1 - 10 | 4.5 | 4.0 |
| Spacer-ext | 18 - 23 | 20 | 4.25 |